Face Identification Using Large Feature Sets

William Robson Schwartz, *Member, IEEE*, Huimin Guo, *Student Member, IEEE*, Jonghyun Choi, *Student Member, IEEE*, and Larry S. Davis, *Fellow, IEEE*

Abstract—With the goal of matching unknown faces against a gallery of known people, the face identification task has been studied for several decades. There are very accurate techniques to perform face identification in controlled environments, particularly when large numbers of samples are available for each face. However, face identification under uncontrolled environments or with a lack of training data is still an unsolved problem. We employ a large and rich set of feature descriptors (with more than 70 000 descriptors) for face identification using partial least squares to perform multichannel feature weighting. Then, we extend the method to a tree-based discriminative structure to reduce the time required to evaluate probe samples. The method is evaluated on Facial Recognition Technology (FERET) and Face Recognition Grand Challenge (FRGC) data sets. Experiments show that our identification method outperforms current state-of-the-art results, particularly for identifying faces acquired across varying conditions.

Index Terms—Face identification, feature combination, feature selection, partial least squares (PLS).

I. INTRODUCTION

FACE recognition has become a very active research area in recent years, mainly driven by its broad applications such as in public security, human–computer interaction, and financial security. The two primary face recognition tasks are *identification* and *verification*. In the identification task, an image of an unknown person is matched to a gallery of known people. In verification, the task is to accept or deny the identity claimed by a person. Therefore, given two face images, the goal is to decide whether the two images are of the same individual or not. The method described in this paper addresses the identification task.

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 [1], [2]. However, when this task is performed under uncontrolled conditions such as uncontrolled

lighting and changes in facial expressions, recognition rates significantly decrease. Face appearances may change when acquisition conditions are uncontrolled, making the recognition problem harder. For example, there can be some extreme illuminations, expressions, and out-of-focus images.

Due to the large size of realistic galleries, not only the accuracy but also the scalability of a face identification system needs to be considered. The main scalability issues are the following. First, the number of subjects in the gallery can be quite large; hence, common search techniques such as brute force nearest neighbor employed to match probe faces do not scale well. Second, in applications such as surveillance and human–computer interaction, in which new subjects are incrementally added, the necessity of rebuilding the gallery models every time a new subject is added compromises the computational performance of the system.

Focusing on the benefits provided by combining an increasing number of feature descriptors weighted by partial least squares (PLS) to emphasize those that best discriminate among different subjects, this paper focuses on the robustness under uncontrolled acquired images and scalability in large galleries.

In order to reduce the problems associated with data collected under uncontrolled conditions, we consider a combination of low-level feature descriptors based on different clues (such approaches have provided significant improvements in object detection [3], [4] and recognition [5], [6]). Then, feature weighting is performed by PLS, which handles very high-dimensional data presenting multicollinearity and works well when very few samples are available [4], [7]–[10]. Finally, a one-against-all classification scheme is used to model the subjects in the gallery.

A downside of employing a one-against-all classification scheme is the imbalance of the class distributions. However, studies have shown that data imbalance presents little influence on the performance of PLS modeling [11], [12]. Barker and Rayens [11] pointed out that PLS involves eigendecomposition of the between-class scatter matrix solely, which only involves calculation of mean vectors of different classes. This does not depend on the number of samples in each class. In addition, Qu *et al.* [12] showed that the weight estimation performed by PLS helps it to extract favorable features for unbalanced classification.

To make the method scalable to the gallery size, we modify the one-against-all approach to use a tree-based structure. At each internal node of the tree, a binary classifier based on PLS regression is used to guide the search for the matching subject in the gallery. The use of this structure provides substantial reduction in the number of comparisons when a probe sample is matched against the gallery.

The main contributions of this paper are the following: 1) the use of PLS regression to weight a combination of a large number of feature descriptors (with more than 70 000 descriptors) that

capture different visual information in a one-against-all classification scheme with highly unbalanced class distributions with a single or very few samples in the positive class; and 2) the application of a tree structure, also based on PLS, to reduce the computational cost of matching probe samples to the gallery.

Our proposed face identification approach outperforms the state-of-the-art techniques in most of the comparisons considering standard face recognition data sets, particularly when data are acquired under uncontrolled conditions such as in experiment 4 of the FRGC data set. In addition, our approach can also handle the problem of insufficient training data, i.e., experimental results show high performance when only a single sample per subject is available. Finally, due to the incorporation of the tree-based structure, a significant number of comparisons can be saved when compared to approaches based on brute force nearest neighbor search.

The rest of this paper is organized as follows. Section II reviews the related work in face identification and feature descriptors. In Section III, we present our face identification method using large feature sets and PLS in a one-against-all classification scheme, and then, we describe a tree-based structure to speed up the identification process. Experimental results with analysis and comparisons with state-of-the-art methods are shown in Section IV. Finally, Section V concludes with some final remarks.

II. RELATED WORK

Face recognition research has achieved significant progress over the past decade. There are some recent and comprehensive surveys on face recognition written by Tolba *et al.* [1] and Zhao *et al.* [2]. Most previous works are based on statistical learning or local matching methods [13]. Methods in the former category use the whole face region to perform recognition and include techniques such as subspace discriminant analysis, support vector machine (SVM), and AdaBoost. These methods suffer from the generalizability problem due to the unpredictable distribution of real-world testing face images, which might be dramatically different from the training samples [14]. Methods in the latter category first locate several facial features and then classify faces according to local statistics.

The use of low-level feature descriptors has been an effective approach in face recognition [5], [14]–[17]. Scale-invariant feature transform (SIFT) and histogram of oriented gradients (HOG), which can be viewed as a quantized code of the facial gradients, are used in face recognition as effective descriptors [18], [19]. Local binary patterns (LBP) and Gabor filters are descriptors most widely used in face recognition. LBP is invariant to monotonic photometric change and can be efficiently extracted. Gabor features are characterized by spatial frequency, spatial locality, and orientational selectivity for coping with image variability such as illumination variations. There are several combinations or variations based on these LBP and Gabor descriptors [5], [14]–[16]. In addition, by varying a sampling radius R and combining the LBP images, a multiresolution representation based on LBP called multiscale LBP (MSLBP) [20] can be obtained. This representation has been suggested

for texture classification, and the results reported for this application show better accuracy than that of the single-scale LBP method. Recent research has focused on parameter learning with a HOG-like template [21], [22]. Other LBP variants, including three-patch-based LBPs, four-patch-based LBPs [23], and Patterns of Oriented Edge Magnitudes [24] have been introduced for face recognition.

Most recently developed face recognition systems work well when images are obtained under controlled conditions or when the test image is captured under similar conditions to those for the training images. However, under varying lighting, aging effects, or blurring images, their performance is still unsatisfactory. To perform recognition under fairly uncontrolled conditions, Tan and Triggs [17] proposed a preprocessing chain for illumination normalization. They used local ternary patterns and a Hausdorff-like distance measure. Holappa *et al.* [25] used LBP texture features and proposed a filter optimization procedure for illumination normalization. Aggarwal *et al.* [26] presented a physical model using Lambert's law to generalize across varying situations. Shih and Liu [27] proposed a new color space LC_1C_2 as a linear transformation of the RGB color space. Using the low-frequency Fourier phase information, Ojansivu *et al.* [28] proposed a texture descriptor called local phase quantization (LPQ), which is insensitive to blurring and rotation. Liu [29] integrates Gabor image representation, a novel multiclass kernel Fisher analysis method, and fractional power polynomial models to improve face recognition performance.

In recent years, face recognition via sparse representation-based classification (SRC) [30] has become very popular. The method uses an $l1$-minimization model, representing a test sample as the sparse combination of the training dictionary. Although SRC performs well when sufficient training images are available, it is still inadequate for many real-world applications where only a single sample per subject is available. Lui and Beveridge [31] proposed the use of Grassmann registration manifolds (GRM), which embed registration images (synthesized images generated by perturbing affine registration parameters of an image) on a Grassmann manifold. They use those perturbed images to form a tangent space and embed the approximated tangent spaces on a Grassmann manifold. A chordal distance is then employed to compare subspaces.

Another challenge is that most current face recognition algorithms perform well when several training images are available per subject; however, they are not adequate for scenarios where a single sample per subject is available. In real-world applications, one training sample per subject presents advantages such as ease of collecting galleries, low cost for storage, and lower computational cost [32]. In [33], Liu *et al.* proposed representing each single (training and testing) image as a subspace spanned by synthesized shifted images and designed a new subspace distance metric. De la Torre *et al.* [34] proposed representational oriented component analysis (ROCA), an extension of oriented component analysis (OCA) [35], to perform face recognition when just one sample per training class is available. ROCA combines several OCA classifiers based on different image representations of the unique training sample.

Regarding the scalability issues previously discussed, there is also previous work focused on scaling recognition systems to large data sets [36]. A technique for combining rejection classifiers into a cascade is proposed by Yuan *et al.* [37] to speed up the nearest neighbor search for face identification. Guo and Zhang [38] proposed the use of a constrained majority voting scheme for AdaBoost to reduce the number of comparisons needed.

III. PROPOSED METHOD

Here, we first describe the feature extraction process and a brief review of PLS regression. Then, the proposed face identification approach is explained in two steps. Initially, we describe the one-against-all approach, then we describe the tree-based structure, which improves scalability when the gallery is large and reduces the computational cost of matching probe samples.

A. Feature Extraction

After cropping and resizing the faces, each sample is decomposed into overlapping blocks, and then, a set of low-level feature descriptors is extracted from each block. The feature extraction methods used capture information related to shape (HOG [39]), texture (captured by LBP [15] and MSLBP [20]), color information (captured by averaging the intensities of pixels in a block, referred to as mean feature), and salient visual properties (captured by Gabor filters [40]).

HOG captures edge or gradient structures that are characteristic of local shape. According to Dalal and Triggs [39], a consequence is a controllable degree of invariance to local geometric transformations, providing invariance, for example, to translations and rotations smaller than the local spatial or orientation bin size.

LBP characterizes the spatial structure of the local image texture and is invariant to monotonic transformations of the pixel gray values [15]. Its original version labels the pixels of an image by thresholding the 3×3 neighborhood with intensity g_p ($p = 0, 1, 2, \ldots, 7$) with respect to its intensity of the center pixel g_c, then defines

$$S(g_p - g_c) = \begin{cases} 1, & g_p \geq g_c \\ 0, & g_p < g_c. \end{cases} \quad (1)$$

Then, the LBP pattern of the image neighborhood is obtained by summing the corresponding thresholded values $S(g_p - g_c)$ weighted by a binomial factor of 2^p as

$$\text{LBP} = \sum_{p=0}^{7} S(g_p - g_c) 2^p. \quad (2)$$

Finally, a 256-bin histogram of the resulting labels is used as a feature descriptor for a patch of the image.

According to the size of the neighborhood employed, there are different versions of LBP. The 3×3 version described above is denoted as LBP_8 due to the use of eight adjacent pixels radially spaced by $45°$. LBP can be also employed in a multiresolution framework by considering concentric circles of different radii called MSLBP [20]. This method has not been widely used in face recognition.

In this paper, in addition to the basic single-scale LBP operator, we consider the MSLBP with setup $\text{LBP}_{8,2}$ (8 pixels on a circle whose radius is 2 pixels) and $\text{LBP}_{8,4}$ (8 pixels on a circle whose radius is 4 pixels). The two resulting histograms are simply concatenated and used as descriptors.

Gabor filters are widely used in object recognition since they capture a number of salient visual properties, including spatial localization, orientation selectivity, and spatial frequency selectivity, quite well [40]. They are robust to illumination variations since they detect amplitude-invariant spatial frequencies of pixel gray values. Gabor filters most commonly used in face recognition have the form

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{\left(-\|k_{\mu,\nu}\|^2 \|z\|^2 / 2\sigma^2\right)} \left[e^{ik_{\mu\nu}z} - e^{-\sigma^2/2}\right] \quad (3)$$

where μ and ν define the orientation and scale of the Gabor kernels $z = (x, y)$, $\| \cdot \|$ denotes the norm operator, and the wave vector is $k_{\mu,\nu} = k_\nu(\cos \phi_\mu, \sin \phi_\mu)$ where $k_\nu = k_{\max}/f^\nu$ and $\phi_\mu = \pi\mu/8$ with $k_{\max}$ being the maximum frequency and f being the spacing factor between kernels in the frequency domain. In this paper, we used $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$.

The Gabor representation of a face is derived from convolving the grayscale face image with the Gabor filters. Let $I(x, y)$ be the face image; its convolution with a Gabor filter is defined as follows:

$$G_{\psi I}(x, y, \mu, \nu) = I(x, y) * \psi_{\mu\nu}(z) \quad (4)$$

where $*$ denotes the convolution operator. Five scales $\mu \in \{0, \ldots, 4\}$ and eight orientations $\nu \in \{0, \ldots, 7\}$ are used here, which results in 40 Gabor filters. For each Gabor filter, one magnitude is computed at each pixel position, resulting therefore in 40 descriptors per pixel. Then, the final feature vector is obtained by downsampling the Gabor features by a factor 4 (one per four rows and columns) in order to reduce the dimensionality of the feature vector to manageable sizes.

After the feature extraction process is performed for all blocks inside a cropped face, descriptors are concatenated, creating a high-dimensional feature vector $\boldsymbol{v}$. This vector is used to describe the face.

B. PLS Regression

PLS is a method for modeling relations between sets of observed variables by means of latent variables. The basic idea of PLS is to construct new predictor variables, i.e., latent variables, as linear combinations of the original variables summarized in a matrix $\boldsymbol{X}$ of descriptor variables (features) and a vector $\boldsymbol{y}$ of response variables. Detailed description of the PLS method can be found in [7], [41], and [42].

Let $\mathcal{X} \subset \mathbb{R}^m$ denote an m-dimensional feature space, and let $\mathcal{Y} \subset \mathbb{R}$ be a scalar space representing the response variable. Let the number of samples be n. PLS decomposes a mean-centered matrix $\boldsymbol{X}_{n \times m}$ and a mean-centered vector $\boldsymbol{y}_{n \times 1}$ 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 PLS (NIPALS) algorithm [7], a set of weight vectors is constructed, stored in matrix $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p)$, such that

$$[\text{cov}(\boldsymbol{t}_i, \boldsymbol{u}_i)]^2 = \max_{|\boldsymbol{w}_i|=1} [\text{cov}(\boldsymbol{X}\boldsymbol{w}_i, \boldsymbol{y})]^2 \quad (5)$$
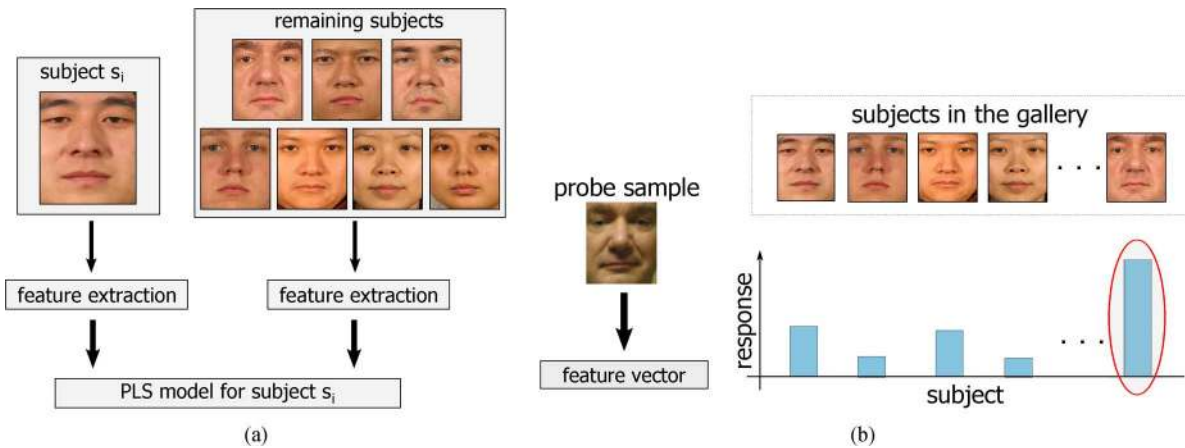
Fig. 1. One-against-all face identification approach. (a) Construction of the PLS regression model for a subject in the gallery. (b) Matching of a probe sample against the subjects in the gallery. The best match for a given probe sample is the one associated with the PLS model presenting the highest regression response.

where $|\boldsymbol{w}_i|$ denotes the 2-norm of vector $\boldsymbol{w}_i$, $\boldsymbol{t}_i$ is the ith column of matrix $\boldsymbol{T}$, $\boldsymbol{u}_i$ is the ith column of matrix $\boldsymbol{U}$, and $\mathrm{cov}(\boldsymbol{t}_i, \boldsymbol{u}_i)$ is the sample covariance between latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. After extracting latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$, matrix $\boldsymbol{X}$ and vector $\boldsymbol{y}$ are deflated by subtracting their rank-one approximations based on $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. This process is repeated until the desired number of latent vectors has been extracted.

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

$$\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^T \boldsymbol{W})^{-1} \boldsymbol{T}^T \boldsymbol{y}. \tag{6}$$

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

$$y_v = \overline{y} + \boldsymbol{\beta}^T \boldsymbol{v} \tag{7}$$

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

It is important to point out that, although the number of weight vectors used to create the low-dimensional representation of data matrix $\boldsymbol{X}$ is p, (7) shows that only a single dot product of a feature vector with the regression coefficients is needed to obtain the response of a PLS regression model, and it is this response that is used to rank faces in a gallery. This characteristic makes the use of PLS particularly fast for finding matches for probe samples, in contrast to other methods where the number of dot product evaluations depends on the number of eigenvectors considered, which is quite large in general [43].

C. One-Against-All Approach

The procedure to learn models for N subjects in the gallery $g = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$, where $\boldsymbol{s}_i$ represents exemplars of each subject's face, is illustrated in Fig. 1(a) and described in detail as follows. Each $\boldsymbol{s}_i$ is composed of feature vectors extracted from cropped faces containing examples of the ith subject.

Since face identification is a multiclass problem, a one-against-all scheme is used to learn a PLS regression model for each subject in the gallery. Therefore, when the ith subject is considered, the remaining samples $g \setminus \boldsymbol{s}_i$ are used as counterexamples of the ith subject. PLS will automatically give higher weights to more discriminatory features when

building each model. To build the PLS model, samples belonging to the ith subject will have response variable $y = +1$ and counterexamples will have $y = -1$. In addition, if the face data set provides a training set, we also add those samples (excluding samples from the subject under consideration) as counterexamples of the ith subject. Experiments show that the addition of training samples as counterexamples improves recognition rates.

After all N models are learned in the gallery, PLS regression models are stored to be used later to evaluate the responses for a probe sample. When a probe image is presented, its feature vector is first extracted and then projected onto each one of these N models. The best match for the probe image is chosen to be the one associated with the PLS model presenting the highest response (due to the scheme chosen to label examples and counterexamples, a higher regression response for the ith PLS model indicates that the probe sample is more similar to the subject associated with this model), as illustrated in Fig. 1(b). Algorithm 1 summarizes both the learning and matching procedures for the one-against-all scheme employed.

Algorithm 1: One-against-all approach.

Learning:

extract features for all subjects

for $i = 1$ *to* N **do**

    set samples of the ith subject as positive exemplars

    set remaining samples as counterexamples

    build the ith PLS regression model

    save regression coefficients $\boldsymbol{\beta}_i$

end

Matching of probe samples:

foreach *probe sample* r **do**

    $\boldsymbol{v} \leftarrow$ extract features from r

    for $i = 1$ *to* N **do**

        $y_v^i \leftarrow \boldsymbol{\beta}_i^T \boldsymbol{v}$

    end

    set rank-1 match of r to be subject $\arg\max_i(y_v^i)$
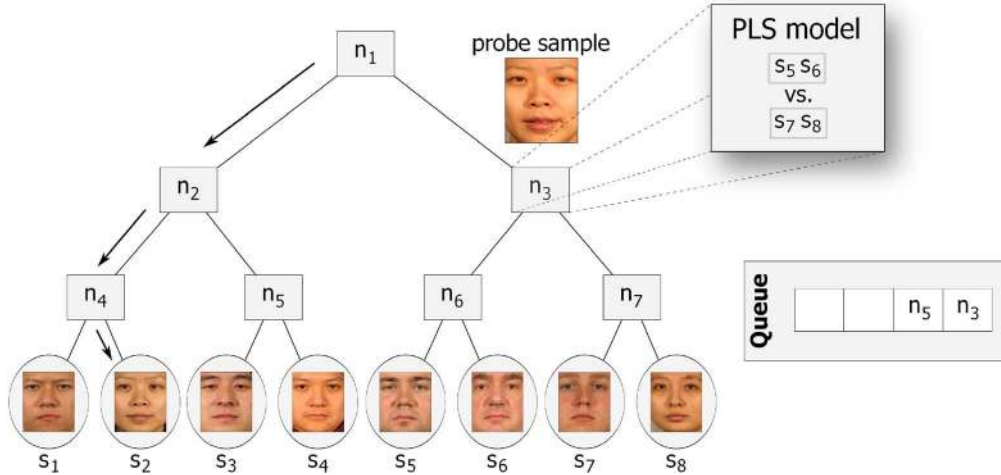
end

Fig. 2. Tree-based structure used to optimize the search for matches to a probe sample. Each internal node contains a PLS regression model used to guide the search, as shown in detail for node n_3, which has a PLS model constructed so that the response directs the search either to node n_6 or node n_7. In this example, the first path to be traversed is indicated by arrows (in this case, it leads to the correct match for this particular probe sample). Alternative search paths are obtained by adding nodes that have not been visited into a priority queue (in this example, nodes n_3 and n_5 will be the starting nodes for additional search paths). After pursuing a number of search paths leading to different leaf nodes, the best match is chosen to be the one presenting the highest response (in absolute value).

D. Optimization Based on a Tree-Based Structure

In terms of scalability, the one-against-all classification scheme described in the previous section has the drawback that, in order to find the best match to a probe sample, the feature vector representing this sample needs to be projected onto all PLS models learned for the subjects in the gallery (common problem faced by methods that estimate matching scores using brute force nearest neighbor search [37]).

To reduce the need for projecting features onto all PLS models to find the best match for a probe sample, we construct a binary tree in which each node n_j contains a subset of the gallery subjects $t_j \subset g$, where $g = \{s_1, s_2, \ldots, s_N\}$, as previously defined. A splitting procedure is used to decide which elements of t_j will belong to the left and right children of n_j, assigning at least one sample to each child. Each internal node is associated with a PLS regression model, used afterward to guide the search when probe samples are analyzed. In order to build the regression model for a node, the subjects assigned to the left child are defined to have response -1 and the subjects assigned to the right child are defined to have response $+1$. The splitting procedure and the building of PLS models are recursively applied in the tree until a node contains only a single subject (leaf node).

The application of the described procedure for a gallery with n subjects results in a tree containing n leaf nodes and $n-1$ PLS regression models located on the internal nodes.

We consider two approaches to split subjects between the children nodes. First, a procedure that uses principal component analysis (PCA) to create a low-dimensional subspace (learned using samples from a training set) and then the K-means algorithm clusters data into two groups, each one is assigned to one child. The second approach chooses random splits and divides the subjects equally into two groups. We evaluate these splitting procedures in Section IV-C .

When a feature vector describing a probe sample is analyzed to find its best matching subject in the gallery, a search starting from the root of the tree is performed. At each internal node, the feature vector is projected onto the PLS model, and according to its response, the search continues either from the left or from the right child. The search stops when a leaf node is reached. Fig. 2 illustrates this procedure.

According to the experimental results shown in Section IV-C, the traversal of a few search paths is enough to obtain the best match for a probe sample. Starting nodes for alternative search paths are stored in a priority queue. An internal node n_k is pushed into the priority queue when its sibling is chosen to be in the current search path. The priority associated with n_k is proportional to its response returned by the PLS regression model at its parent. Finally, since each search path leads to a leaf node, the best match for a given probe sample is chosen to be the one presenting the highest response (in absolute value) among the leaf nodes reached during the search.

IV. EXPERIMENTS

Here, we evaluate several aspects of our proposed approach. Initially, we show that the use of the low-level feature descriptors analyzed by PLS in a one-against-all scheme, as described in Section III-C, improves recognition rates over previous approaches, particularly when the data are acquired under uncontrolled conditions (changes in illumination, expression, and aging effects). Then, we demonstrate that the tree-based approach introduced in Section III-D obtains comparably high recognition rates with a significant reduction in the number of projections.

The method is evaluated on two standard data sets used for face recognition, namely, Facial Recognition Technology (FERET) and Face Recognition Grand Challenge (FRGC) version 1. The main characteristics of the FERET data set are that it contains a large number of subjects (single image per subject) in the gallery and the probe sets exploit differences in illumination, facial expression variations, and aging effects [44]. Since FRGC contains faces acquired under uncontrolled conditions [45], we evaluate the influence of several aspects of

the proposed method, such as the addition of training samples as counterexamples on the one-against-all classification scheme, image size for face samples, feature channel combination, and the use of the multiclass PLS.

We use annotations regarding eyes, nose, and mouth locations provided by both FERET and FRGC to crop and align facial samples. Although such process makes the direct application of the method less suitable to general images, the results shown allow us to focus only on the discriminative representations being proposed without interference from problems caused by detection errors or misalignment of faces. Therefore, we are able to make a fair comparison with other methods, which also take advantage of the annotations provided.

All experiments were conducted on an Intel Core i7-860 processor, 2.8 GHz with 4 GB of RAM running Windows 7 operating system using a single processor core. The method was implemented using C++ programming language.

A. Evaluation on the FERET Data set

The frontal faces in the FERET database are divided into five sets, i.e., 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; hence, some facial features have changed. The images are cropped and rescaled to 110×110 pixels.

Experimental Setup: Since the FERET data set is taken under varying illumination conditions, we preprocessed the images for illumination normalization. Among the best known illumination normalization methods are the self-quotient image (SQI) [46], total variation models, and anisotropic smoothing [25]. SQI is a retinex-based method, which does not require training images and has relatively low computational complexity; we use it due to its simplicity.

Once the images are normalized, we perform feature extraction. For HOG features, we use block sizes of 32×32 and 16×16 with strides of 8 and 4 pixels, respectively. For LBP features, we use block size of 32×32 with a stride of 16 pixels. The mean features are computed from block size of 4×4 with a stride of 2 pixels. The Gabor features have five scales and eight orientations, downsampled by a factor of 4. The MSLBP uses block size of 32×32 with a stride of 16 pixels at radii 2 and 4. The final length of the concatenated features is 74 724. Experiments performed with a number of weight vectors for the PLS model varying from 13 to 21 provided very similar results; therefore, we have chosen the number of weight vectors $p = 19$ for all experiments with the FERET data set.

Results and Comparisons: Fig. 3 shows the cumulative match characteristic curves obtained by the one-against-all approach described in this paper, referred to as *extended PLS*, for all FERET probe sets. We see that our method is robust to facial expressions (fb), lighting (fc), and aging effect ($dup1, dup2$).

Table I shows the rank-1 recognition rates of previously published algorithms and ours on the FERET data set. The results
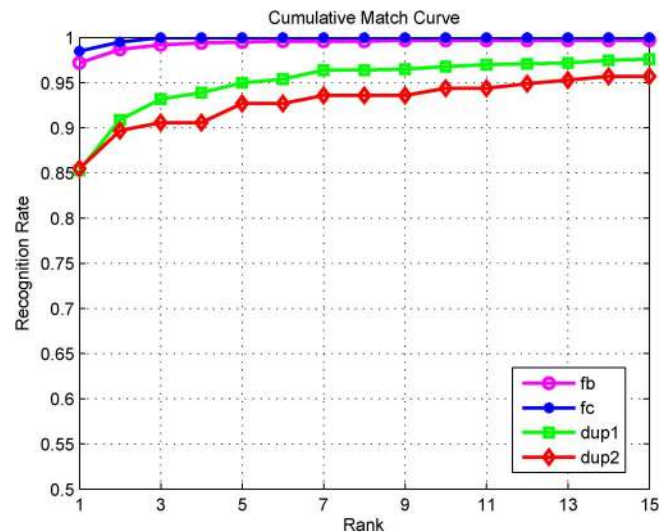

Fig. 3. Cumulative score of the top 15 matches of our identification algorithm for the FERET probe sets.

TABLE I
RECOGNITION RATES OF OUR IDENTIFICATION ALGORITHM (REFERRED TO
AS EXTENDED PLS) AND COMPARISONS WITH OTHER
ALGORITHMS FOR THE FERET PROBE SETS

	Method	fb	fc	dup1	dup2
use training set	Best of [44]	95	82	59	52
	LBP [15]	97	79	66	64
	Tan [5]	**98**	98	**90**	85
not use training set	LGBPHS [14]	**98**	97	74	71
	HGPP [16]	97.6	98.9	77.7	76.1
	SIS [33]	91	90	68	68
	GRM-H [31]	94.1	92.8	70.8	76.9
	GRM-L [31]	97.5	97.9	79.5	83.8
	PLS [47]	95.7	**99.0**	80.3	80.3
	extended PLS	97.2	98.5	85.3	**85.5**

of these algorithms can be divided into two groups, i.e., those using a training set and those not using a training set.[1] The training set is commonly used to build a subspace to obtain a low-dimensional representation of the features before performing the match. This subspace provides additional information regarding the domain of the problem. As shown in the table, our one-against-all approach, which does not use a training set, achieves similar results on fb and fc and obtains rank-1 recognition rates higher than 85% on the challenging $dup1$ and $dup2$ sets.

B. Evaluation on the FRGC Data set

We evaluate our method using three experiments from FRGC version 1 that consider 2-D images. Experiment 1 contains a single controlled gallery image and a probe with one controlled still image per subject (183 training images, 152 gallery images, and 608 probe images). Experiment 2 considers the identification of a person given a gallery with four controlled still images per subject (732 training images, 608 gallery images, and 2 432 probe images). Finally, experiment 4 considers a gallery with

---

[1]Different from the pattern recognition context, where the training set consists of patterns of the same classes as the test set, the set referred to as training set in face identification contains samples that might not belong to subjects in the gallery or probe (test set).
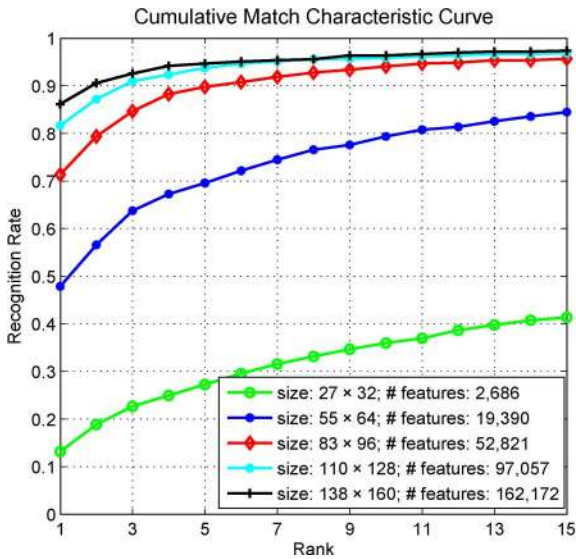
Fig. 4. Recognition rates obtained with different image sizes for experiment 4 of the FRGC data set.
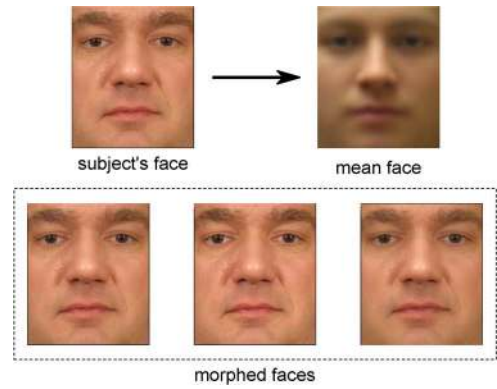

Fig. 5. Original gallery image, mean face, and three morphed images used to learn the PLS regression model.

Fig. 6. Recognition rates obtained by each feature channel individually and its combination considering experiment 4 of the FRGC data set.

one controlled still image per subject and multiple uncontrolled probe images 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 138×160 pixels.

Experimental Setup: For HOG features, we use block sizes of 32×32 and 16×16 with strides of 8 and 4 pixels, respectively. For LBP features, we use block size of 32×32 with a stride of 16 pixels. In addition, the mean features are extracted from block size of 8×8 with a stride of 4 pixels. The Gabor features have five scales and eight orientations, downsampled by a factor of 4. The MSLBP uses block size of 32×32 with a stride of 16 pixels at radii 2 and 4. The final length of the concatenated features is 162 172. Similar to the FERET data set, the experiments indicated that the number of weight vectors varying from 13 to 21 provided similar results; therefore, we have chosen $p = 13$ for all experiments with the FRGC data set.

Experiment 4 in FRGC version 1 is considered the most challenging in this data set. 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 [48]. 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%, and 40%. This generates three synthesized images. Therefore, for each subject in the gallery, we now have four samples. Fig. 5 depicts three morphed faces from the original image from the gallery. We can see that the illumination is slightly changed and the samples become smoother.

Image Size: To show the influence of image size, which is directly proportional to the number of feature descriptors considered, we considered different sizes for the input samples for experiment 4. The resulting cumulative match characteristic curves are shown in Fig. 4. Improvements in recognition rates are obtained with the use of more feature descriptors
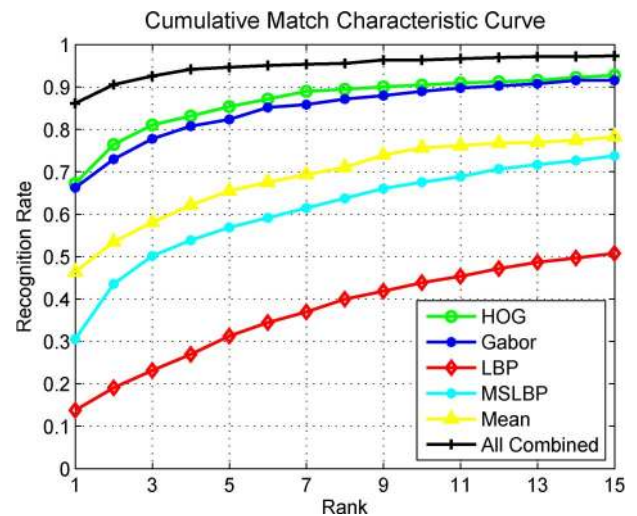
(extracted when larger input images are used). The best results are obtained with images of 138×160 pixels, which results in 162 172 feature descriptors. These results show that extraction of a large amount of information provides a sample description possessing high discriminative power and corroborates that PLS is able to perform feature weighting for large numbers of features and small numbers of samples per class.

Combination of Feature Channels: An important aspect to be evaluated when multiple feature channels are considered is the contribution that each feature channel provides. Fig. 6 shows recognition rates obtained for each feature channel individually and their combination, considering experiment 4. When individually evaluated, HOG provides the best recognition rates and LBP provides the worst ones. However, although HOG achieved satisfactory results, the combination of all features increased recognition results by more than 25%, going from 67.3% to 86.2%. This shows that the features used are nonredundant and their combination provides a better description for different subjects.

According to the results achieved with individual feature channels, it is possible to conclude that features based on shape provide the most information for face recognition. On the other hand, texture features do not provide high performance,
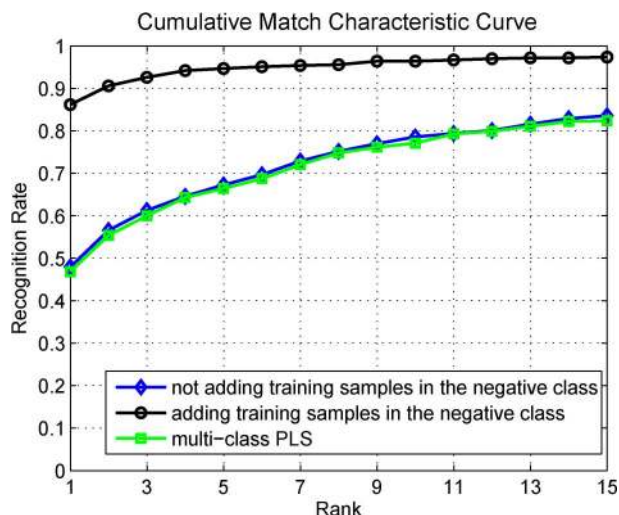
Fig. 7. Addition of training samples as counterexamples in experiment 4 of the FRGC data set and use of the multiclass PLS instead of the one-against-all classification scheme.

indicating that texture is not a strongly discriminative feature for recognition. Finally, although the mean feature is a very simple descriptor based on color, it provides better information than texture.

The results shown in Fig. 6 were obtained using the morphed images, as described in the experimental setup. It is important to note that the use of such images improves the recognition rates. For example, in experiment 4, the rank-1 recognition rate achieved without the morphed images was 80.9%.

Addition of Counterexamples and Multiclass PLS: Fig. 7 shows the improvements obtained when samples from a training set are added as counterexamples for experiment 4. The added samples are not from any subject in the gallery, and half of them are acquired under uncontrolled conditions. Results show that the addition of counterexamples increased recognition rates significantly.

Since the gallery for experiment 4 contains only samples acquired under controlled conditions and probe samples were acquired under uncontrolled conditions, the PLS models built without additional samples do not possess information regarding the environment in which the system will be tested. Therefore, feature descriptors presenting high weight might in fact not be the ones with high discriminative power. The addition of uncontrolled counterexamples prevents such descriptors being chosen as the important ones.

Fig. 7 also shows recognition rates achieved when PLS considers a multiclass problem instead of the one-against-all classification scheme. In this case, vector y (as defined in Section III-B) is changed to a matrix with the number of columns equal to the number of classes. We see that the recognition rates reported are significantly lower compared with the ones obtained by the one-against-all classification scheme with the addition of counterexamples.

Results and Comparisons: Fig. 8 shows the cumulative match characteristic curves obtained by the one-against-all approach for the three probe sets of FRGC, and Table II shows the rank-1 recognition rates of different algorithms on the FRGC probe sets. Our method outperforms others in every probe set
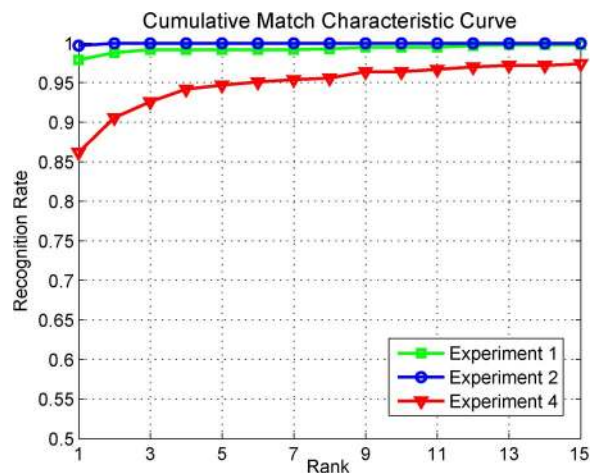

Fig. 8. Cumulative score of the top 15 matches of our identification algorithm for the FRGC probe sets.

TABLE II
RECOGNITION RATES OF OUR IDENTIFICATION ALGORITHM (REFERRED TO AS EXTENDED PLS) AND COMPARISONS WITH OTHER ALGORITHMS FOR THE FRGC PROBE SETS

Method	Exp.1	Exp.2	Exp.4
PCA [50]	87.6	95.6	-
UMD [26]	94.2	99.3	-
BEE (from [27])	-	-	37.0
LC$_1$C$_2$ [27]	-	-	75.0
Tan(from [25])	-	-	58.1
Holappa [25]	-	-	63.7
LPQ [28]	-	-	74.5
LIU [29]	-	-	78
ROCA [34]	96.4	-	75.5
PLS [47]	97.5	99.4	78.2
extended PLS	**97.9**	**99.8**	**86.2**

considered, particularly on the most challenging experiment 4. This is, to the best of our knowledge, the best performance reported in the literature.

In addition to the comparison with other methods, as shown in Table II, we also compared the use of PLS with SVM for experiment 4 of the FRGC data set with the same setup and features. Using linear SVM implemented by libSVM [49], the multiclass SVM achieved rank-1 recognition rate of 21.21% and the one-against-all setup achieved 65.95%, as compared to 46.88% and 86.2% achieved using PLS, respectively.

C. Evaluation of the Tree-Based Structure

Here, we evaluate the tree-based structure described in Section III-D. First, we evaluate procedures used to split the set of subjects belonging to a node. Second, we test heuristics used to reduce the search space. Third, we compare the results previously obtained by the one-against-all approach to the results obtained when the tree-based structure is incorporated. Finally, we compare our method to the approach proposed by Yuan *et al.* [37].

To evaluate the reduction in the number of comparisons, here, the x axis of the plots no longer displays the rank; instead, it shows either the percentage of projections performed by the tree-based approach when compared with the one-against-all
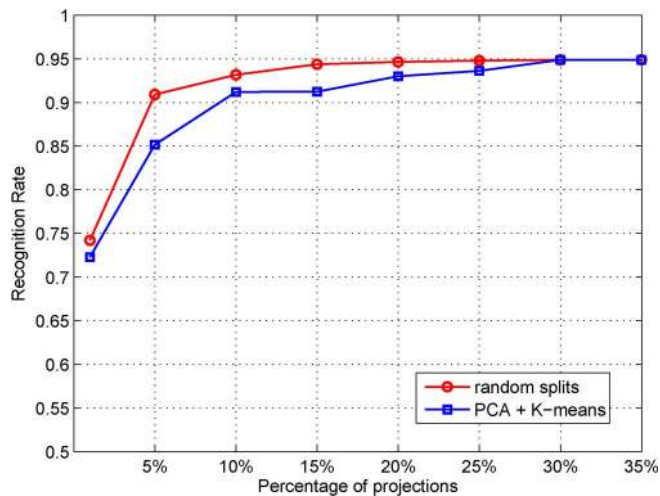
Fig. 9. Comparison of the recognition rates when random splits and PCA + K-means approach are used.

Fig. 10. Evaluation of the heuristic based on stopping the search after a maximum number of tree traversals is reached.

approach (e.g., Fig. 9) or the percentage of tree traversals when compared to the number of subjects in the gallery (e.g., Fig. 10). The y axis displays the recognition rates for the rank-1 matches. We used the probe set fb from the FERET data set to perform evaluations here.

Procedure to Split Nodes: Fig. 9 shows that both splitting procedures described in Section III-D obtain similar recognition rates when the same number of projections is performed. Due to the similarity of the results, we have chosen to split the nodes randomly. The advantages of applying random splits are the lower computational cost to build the gallery models and balanced trees are obtained. Balanced trees are important since the depth of a leaf node is proportional to $\log_2 N$, which is desirable to keep short search paths. For this experiment, the number of tree traversals was set to 15% and the heuristic based on the median was not considered (these heuristics will be discussed in the next paragraphs).

Heuristics to Reduce the Search Space: The first experiment evaluates the recognition rate as a function of the maximum number of traversals allowed to find the match subject to a probe sample; this is limited to a percentage of the gallery size (a traversal is defined as the path between a starting node, which is obtained from the priority queue, and a leaf node). Fig. 10 shows the maximum recognition rates achievable for a given percentage. We can see that as low as 15% of traversals are enough to obtain recognition rates comparable with the results obtained by the one-against-all approach (95.7% for the probe set considered in this experiment). Therefore, in the remaining experiments, we limit the number of traversals to 15% of the gallery size.

In the second experiment, we consider the following heuristic. For the initial few probe samples (15 samples), all search paths are evaluated and the absolute values of the regression responses for the best matches are stored. The median of these values is computed. Then, for the remaining probe samples, the search is stopped when the regression response for a leaf node is higher than the estimated median value (and
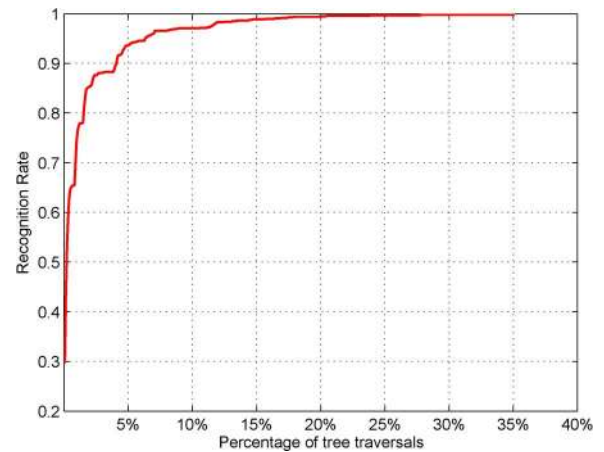
the subject represented at that leaf node is set to be the best match for the probe sample). Our experiments show that this heuristic alone is able to reduce the number of projections to 63% without any degradation in recognition rates.

Results and Comparisons: Using the results obtained from the previous experiments (random splits and adding both heuristics to reduce the search space), we now compare the recognition rates obtained when the tree-based structure is used to the results obtained by the one-against-all approach. Then, we evaluate the speedup achieved by reducing the number of projections.

Fig. 11(a) and (b) shows the identification results obtained for FERET and FRGC data sets, respectively. Overall, we see that, when the number of projections required by the one-against-all approach is reduced to 20% or 25%, there is a negligible drop in the recognition rate, as shown in the previous sections. Therefore, without decreasing the recognition rate, the use of the tree-based structure provides a clear speedup for performing the evaluation of the probe set. According to the plots, speedups of 6.66 times are achieved for FERET, and for FRGC, the speedup is up to 10 times depending on the experiment being considered. Furthermore, since the number of projections is equal to the number of internal nodes, only a small number of the nodes are visited.

In addition to the speedup achieved to match a probe sample to the gallery, the tree-based structure can be also exploited to facilitate incremental enrollment, avoiding rebuilding all PLS models. Due to its hierarchical structure and independence between subjects in different subtrees, when there is a subject enrollment, the only models that need to be rebuilt are those on the path between the root and the leaf node where the new subject is added.

Finally, we compare our method to the *cascade of rejection classifiers* (CRC) approach proposed by Yuan *et al.* [37]. Table III shows the speedups over the brute force nearest neighbor search and rank-1 error rates obtained by both approaches. We apply the same protocol used in [37] for the FRGC data set. Higher speedups are obtained by our method and, different from CRC, no increase in the error rates is noticed when larger test set sizes are considered.
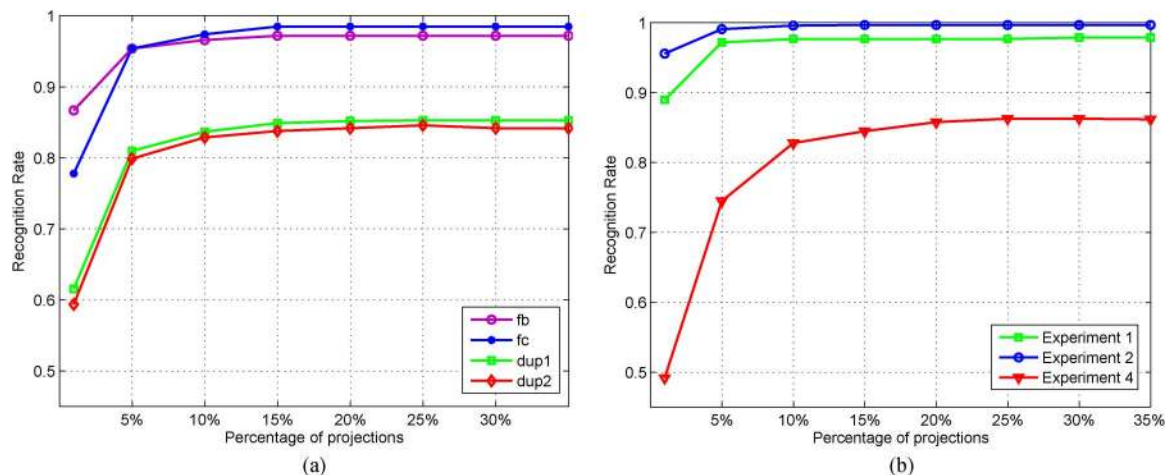
Fig. 11. Recognition rates as a function of the percentage of projections performed by the tree-based approach when compared to the one-against-all approach. (a) FERET. (b) FRGC.

TABLE III
COMPARISON BETWEEN OUR TREE-BASED APPROACH AND THE CRC APPROACH

		test set size as fraction of dataset				
		10%	21%	32%	43%	65%
CRC	speed-up	1.58	1.58	1.60	2.38	3.35
	rank-1 error rate	19.5%	22.3%	24.3%	28.7%	42.0%
Proposed approach	speed-up	3.68	3.64	3.73	3.72	3.80
	rank-1 error rate	5.62%	5.08%	5.70%	5.54%	5.54%

V. CONCLUSION

We have proposed a face identification method using a set of low-level feature descriptors analyzed by PLS, which presents the advantages of being both robust and scalable. Experimental results have shown that the method works well for single image per sample, in large galleries, and under different conditions.

The use of PLS regression makes the evaluation of probe–gallery samples efficient due to the necessity of only a single dot product evaluation. Optimization is further employed by incorporating the tree-based structure, which largely reduces the number of projections when compared with the one-against-all approach, with a negligible effect on recognition rates. In addition, experiments had shown that the combination of multiple feature channels, the use larger image size, and the addition of counterexamples improve results greatly.

ACKNOWLEDGMENT

REFERENCES

[1] A. Tolba, A. El-Baz, and A. El-Harby, "Face recognition: A literature review," *Int. J. Signal Process.*, vol. 2, no. 2, pp. 88–103, 2006.
[2] 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, Dec. 2003.
[3] B. Wu and R. Nevatia, "Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.
[4] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 24–31.
[5] X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2007, pp. 235–249.
[6] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1978–1990, Oct. 2011.
[7] H. Wold, "Partial least squares," in *Encyclopedia of Statistical Sciences*, S. Kotz and N. Johnson, Eds. New York: Wiley, 1985, pp. 581–591.
[8] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proc. Brazilian Symp. Comput. Graph. Image Process.*, 2009, pp. 322–329.
[9] C. Dhanjal, S. Gunn, and J. Shawe-Taylor, "Efficient sparse kernel feature extraction based on partial least squares," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 8, pp. 1347–1361, Aug. 2009.
[10] Q.-S. Sun, Z. Jin, P.-A. Heng, and D.-S. Xia, "A novel feature fusion method based on partial least squares regression," in *Proc. Int. Conf. Adv. Pattern Recog.*, 2005, pp. 268–277.
[11] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemometrics*, vol. 17, no. 3, pp. 166–173, Mar. 2003.
[12] H.-N. Qu, G.-Z. Li, and W.-S. Xu, "An asymmetric classifier based on partial least squares," *Pattern Recognit.*, vol. 43, no. 10, pp. 3448–3457, Oct. 2010.
[13] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2617–2628, Oct. 2007.
[14] 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 *Proc. Int. Conf. Comput. Vis.*, 2005, pp. 786–791.
[15] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 469–481.
[16] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 57–68, Jan. 2007.
[17] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[18] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B.-L. Lu, "Person-specific SIFT features for face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2007, pp. 593–596.

[19] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG-EBGM," *Pattern Recognit. Lett.*, vol. 29, no. 10, pp. 1537–1543, Jul. 2008.

[20] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[21] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 2707–2714.

[22] N. Pinto and D. Cox, "Beyond simple features: A large-scale feature search approach to unconstrained face recognition," in *Proc. Int. Conf. Autom. Face Gesture Recog.*, 2011, pp. 8–15.

[23] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Proc. Real-Life Images Workshop Eur. Conf. Comput. Vis.*, 2008.

[24] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 313–326.

[25] J. Holappa, T. Ahonen, and M. Pietikainen, "An optimized illumination normalization method for face recognition," in *Proc. IEEE Int. Conf. Biometrics: Theory, Appl. Syst.*, 2008, pp. 1–6.

[26] G. Aggarwal, S. Biswas, and R. Chellappa, "UMD experiments with FRGC data," in *Proc. Comput. Vis. Pattern Recog. Workshop*, 2005, pp. 172–178.

[27] P. Shih and C. Liu, "Evolving effective color features for improving FRGC baseline performance," in *Proc. Comput. Vis. Pattern Recog. Workshop*, 2005, pp. 156–163.

[28] V. Ojansivu, E. Rahtu, and J. Heikkilä, "Rotation invariant local phase quantization for blur insensitive texture analysis," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2008, pp. 1–4.

[29] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 725–737, May 2006.

[30] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[31] Y. Lui and J. Beveridge, "Grassmann registration manifolds for face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 44–57.

[32] X. Tan, S. Chen, Z. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, Sep. 2006.

[33] J. Liu, S. Chen, Z. Zhou, and X. Tan, "Single image subspace for face recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2007, pp. 205–219.

[34] F. De la Torre, R. Gross, S. Baker, and V. Kumar, "Representational oriented component analysis (ROCA) for face recognition with one sample image per training class," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 266–273.

[35] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*. New York: Wiley, 1996.

[36] M. Aly, P. Welinder, M. Munich, and P. Perona, "Scaling object recognition: Benchmark of current state of the art techniques," in *Proc. IEEE Conf. Comput. Vis. Workshop*, 2009, pp. 2117–2124.

[37] Q. Yuan, A. Thangali, and S. Sclaroff, "Face identification by a cascade of rejection classifiers," in *Proc. Comput. Vis. Pattern Recog. Workshop*, 2005, pp. 152–159.

[38] G.-D. Guo and H.-J. Zhang, "Boosting for fast face recognition," in *Proc. Int. Conf. Comput. Vis. Workshop*, 2001, pp. 96–100.

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

[40] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.

[41] L. Elden, "Partial least squares vs. Lanczos bidiagonalization—I: Analysis of a projection method for multiple regression," *Comput. Statist. Data Anal.*, vol. 46, no. 1, pp. 11–31, May 2004.

[42] R. Rosipal and N. Kramer, "Overview and recent advances in partial least squares," in *Proc. SLSFS*, 2006, vol. 3940, Lecture Notes in Computer Science, pp. 34–51.

[43] K. Delac, M. Grgic, and S. Grgic, "Independent comparative study of PCA, ICA, and LDA on the FERET data set," *Int. J. Imaging Syst. Technol.*, vol. 15, no. 5, pp. 252–260, 2005.

[44] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[45] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 947–954.

[46] H. Wang, S. Z. Li, and Y. Wang, "Face recognition under varying lighting conditions using self quotient image," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 819–824.

[47] W. Schwartz, H. Guo, and L. Davis, "A robust and scalable approach to face identification," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 476–489.

[48] B. Kamgar Parsi, E. Lawson, and P. Baker, "Toward a human-like approach to face recognition," in *Proc. IEEE Int. Conf. Biometrics: Theory, Appl. Syst.*, 2007, pp. 1–6.

[49] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.* vol. 2, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[50] A. Mian, M. Bennamoun, and R. Owens, "2D and 3D multimodal hybrid face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 344–355.

William Robson Schwartz (M'11) received the B.Sc. and M.Sc. degrees in computer science from the Federal University of Paraná, Curitiba, Brazil, and the Ph.D. degree in computer science from the University of Maryland, College Park.

He is currently a Postdoctoral Researcher with the Institute of Computing, University of Campinas, Campinas, Brazil. His research interests include computer vision, biometrics, pattern recognition, and image processing.

Huimin Guo (S'10) received the B.S. and M.S. degrees in computer science from Beijing Normal University, Beijing, China. While taking up the Master's degree, she exchanged to Waseda University, Tokyo, Japan, under the 1.5-year Japanese Government [Ministry of Education, Culture, Sports, Science and Technology (MEXT)] Scholarship. She is currently working towards the Ph.D. degree in the Department of Computer Science, University of Maryland, College Park.

Her research focuses on computer vision and pattern recognition.

Jonghyun Choi (S'10) received the B.S. and M.S. degrees in electrical engineering and computer science from Seoul National University, Seoul, Korea, in 2003 and 2008, respectively. He is currently working towards the Ph.D. degree in the Computer Vision Laboratory, University of Maryland, College Park.

His research interests include machine learning, optimization, and image analysis and representation.

Larry S. Davis (F'98) received the B.A. degree from Colgate University, Hamilton, NY, in 1970 and the M.S. and Ph.D. degrees in computer science from the University of Maryland, College Park, in 1974 and 1976, respectively.

From 1977 to 1981, he was an Assistant Professor with the Department of Computer Science, University of Texas, Austin. He returned to the University of Maryland as an Associate Professor in 1981. From 1985 to 1994, he was the Director of the University of Maryland Institute for Advanced Computer Studies, where he is currently a Professor. He is also a Professor and the Chair of Department of Computer Science.