

Classification Schemes based on Partial Least Squares for Face Identification

Gérson de Paulo Carlos^a, Helio Pedrini^a, William Robson Schwartz^{b,c}

^a*Institute of Computing, University of Campinas, Campinas-SP, Brazil, 13083-852*

^b*Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil, 31270-010*

^c*Corresponding author (e-mail: william@dcc.ufmg.br)*

Abstract

Approaches based on the construction of highly discriminative models, such as one-against-all classification schemes, have been employed successfully in face identification. However, their main drawback is the reduction in the scalability once the models for each individual depend on the remaining subjects. Therefore, when new subjects are enrolled, it is necessary to rebuild all models to take into account the new individuals. This work addresses different classification schemes based on Partial Least Squares employed to face identification. First, the one-against-all and the one-against-some classification schemes are described and, based on their drawbacks, a classification scheme referred to as one-against-none is proposed. This novel approach considers face samples that do not belong to subjects in the gallery. Experimental results show that it achieves similar results to the one-against-all and one-against-some even though it does not depend on the remaining subjects in the gallery to build the models.

Keywords: Face identification, partial least squares, one-against-none, one-against-all, one-against-some.

1. Introduction

Face recognition is a very active research topic due to its applications in areas such as surveillance, biometrics and human computer interaction. Face verification and identification figure among the main tasks performed by face recognition. While the former is responsible for accepting or denying the identity claimed by an individual given a pair of samples, the latter focuses on matching a sample of an unknown person to a gallery of known subjects.

The identification task presents particular interest in surveillance applications that perform face recognition in monitored areas, in which the identity of individuals needs to be determined to provide, for example, non-intrusive monitoring of circulation on restricted areas. Due to the dynamic nature of these environments, in which new subjects are incrementally added, the identification system not only needs to be accurate, but also it is important to provide efficient and robust enrollment mechanisms.

Due to its ability of generating discriminative subspaces and working with few high dimensional input samples, the statistical method *Partial Least Squares* (PLS) [1] has been successfully applied to the problem of face recognition in the past few years for both verification and identification tasks [2, 3, 4, 5, 6, 7]. These works employ the one-against-all classification scheme combined with PLS, which provides high recognition rates for the identification task [2, 4, 6]. However, this classification scheme presents the drawback of not being scalable to the enrollment of new subjects since all existing PLS models representing the subjects in the gallery need to be rebuilt, leading to a high computational cost proportional to the gallery size.

To handle the enrollment of new subjects to the gallery while maintain-

ing the generation of highly discriminative subspaces with PLS, we have recently proposed a classification scheme, called one-against-some [8], which does not consider all remaining subjects as counterexamples, but only a subset of them. The one-against-some approach maintains a trade-off between the discriminatory power achieved by the one-against-all, in which all remaining subjects are set as counterexamples, and allows scalability when new subjects are enrolled in the gallery since the PLS models used to represent subjects already enrolled do not need to be rebuilt. However, as a drawback, the number of PLS models that have to be built is larger than the number required by the one-against-all approach and a priority queue is employed to maintain a low number of projections when a probe sample is presented to the system.

In this work, we propose a novel classification scheme based on PLS models, called *one-against-none*, which does not require the addition of samples of the remaining subjects as counterexamples, but only the addition of fixed set of samples that are not required to belong to any subject under consideration. These samples, which do not have intersection with the subject being considered, are referred to as extra samples. When employing the one-against-none classification scheme, the number of PLS models to be built is equal to the number of the subjects, which maintains a low number of projections. In this work, we also evaluate different approaches to choosing the counterexamples in the one-against-some classification scheme.

In the experiments, we conduct an evaluation of the three classification schemes based on PLS: one-against-all, one-against-some and one-against-none, emphasizing their strong and weak points regarding computational cost, memory consumption and accuracy.

2. Related Work

A brief review and discussion of concepts and references related to face recognition, feature descriptors and classification schemes are included in this section.

2.1. Face Recognition

Face recognition has been extensively investigated for decades and is still a challenging problem. Although various approaches have achieved high recognition accuracy rates under specific conditions, several factors can affect their performance such as varying pose, non-uniform illumination, partial occlusion, scalability, facial characteristics, and other uncontrolled conditions.

A face recognition system usually consists of four main components: face detection, alignment, feature extraction, and matching. Some comprehensive surveys on face recognition are available in the literature [9, 10].

Techniques for face recognition are commonly classified into two main categories, local and holistic approaches. Local approaches [11, 12, 13, 14] extract facial features, such as nose, eyes and mouth, to discriminate faces based on combination and evaluation of local statistics, whereas holistic approaches [15, 16, 17, 18] capture information from the entire face image to perform the recognition.

Since the recognition problem can involve a large number of individuals in a gallery, scalability is an important issue in face identification systems. Therefore, search techniques employed to match probe samples in the gallery must be efficient [19, 20]. Additionally, the model reconstruction whenever a new individual is added to the gallery can significantly affect the system performance.

In [21], the authors propose a regularization method of learning a similarity metric for unconstrained face verification. A learning objective function is formulated by considering the discrimination for separating among dissimilar image-pairs and the robustness to the large intrapersonal variations. The performance of their method is evaluated on the Labeled Faces in the Wild (LFW) data set. The work in [22] proposes an unconstrained correlation filter applied to the face recognition problem to extract discriminative features in class-dependence feature analysis (CFA). Experiments are conducted on five data sets (AR, FERET, FRGC, LFW, CAS-PEAL-R1). In [23], the authors propose a correlation filter bank for face recognition, which explores local features and the combination of different face subregions. Experimental results are evaluated on public data sets (FERET, LFW, CAS-PEAL-R1).

Some methods project face images onto a lower dimensional space to reduce data high-dimensionality, such as Eigenface [24], Fisherface [25], Tensorface [26] and Bayesian algorithms [27]. Thus, face variations can be modeled under different illumination conditions through a training face set, where new face samples can be projected onto low dimensional spaces and compared to a number of images. Face recognition methods based on Partial Least Squares (PLS) has recently provided interesting results [2, 3, 4, 5, 6, 7, 8].

2.2. Feature Descriptors

Facial features, such as eyes, nose, mouth and eyebrows, can be extracted by edge-based detectors. However, a common problem with these approaches is that features can be affected by noise, illumination and occlusion. Therefore, several local feature descriptors have been employed

effectively in face description. There are three main approaches, based on texture information to capture local patterns in the face [28, 29, 30, 13], shape [31, 32, 15, 14], and color information [13, 16, 33, 34].

Fusion of descriptors has been explored for face recognition [35, 36], where complementary local and global features extracted from regions and the entire image, respectively, are combined to form a feature vector. To avoid the need for selecting the local descriptors, deep learning approaches have been employed on face recognition so that robust local descriptors can be learned automatically [17].

2.3. Classification Schemes

Some classification strategies can be used in the matching of a probe sample against a face gallery, such as the one-against-all scheme [37], where all training models must be reconstructed when a new subject s_i is presented to the system, since all the remaining samples are employed as counterexamples (negative class) for s_i . Even though this strategy can achieve high recognition rates, it requires high training time.

Pairwise classification [38, 39] converts multi-class problems into a series of two-class (binary) problems. Differently from the one-against-all scheme, which constructs one binary classifier for each class c_i such that positive training samples belong to class c_i and negative training samples are formed by all remaining classes, the pairwise classification converts an n -class problem into $n(n - 1)/2$ binary problems, one for each pair of classes.

One-against-some scheme [8, 40] has been proposed for the person re-identification problem, where only a subset of subjects are considered as negative samples, instead of all remaining individuals as in the one-against-all scheme. Thus, not all individuals need to be known for the enrollment

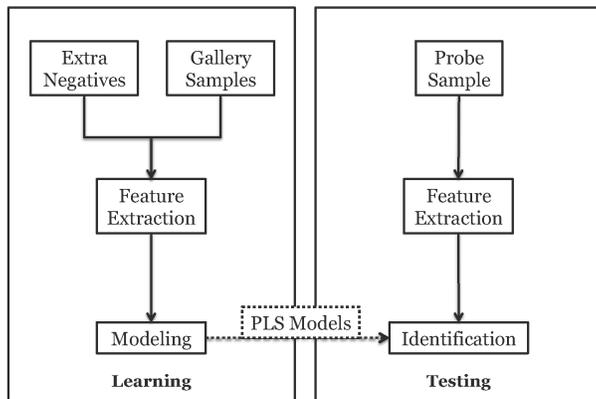


Figure 1: General face identification approach based on Partial Least Squares.

of a new subject, avoiding the reconstruction of the existing gallery models to handle the new individual.

3. Classification Schemes based on Partial Least Squares

In this section, we present the classification schemes based on Partial Least Squares which have been employed to the face identification problem. We describe two previously proposed approaches: the *one-against-all* classification scheme [2] and the *one-against-some* classification scheme [8]. In addition, based on their drawbacks regarding the enrollment of new subjects, we propose the classification scheme called *one-against-none*.

The face identification approaches based on subject modeling and Partial Least Squares are structured in three main blocks: feature extraction, individual modeling for subjects in the gallery and execution of regressions to perform the matching of probe samples to subjects in the gallery. The general approach is illustrated in Figure 1 and each of its parts is described in details in Sections 3.1-3.4, pointing out the differences among the three classification schemes. Finally, their advantages and drawbacks are discussed

and summarized in Section 3.5.

3.1. Partial Least Squares

Partial Least Squares (PLS) aims at finding relations between observed variables through the estimation of a low dimensional latent space to maximize the separation between samples with different characteristics [41] (samples belonging to different subjects in our context). Given a matrix \mathbf{X} , containing feature descriptors extracted from each sample in its rows, and a matrix \mathbf{Y} containing a set of response variables, the PLS approach estimates a set of latent variable as a linear combination of the variables in the matrix \mathbf{X} .

The PLS performs the latent variable estimation as follows. Let $\mathbf{X}_{n \times d}$ be a matrix with n samples described by d variables each, and let $\mathbf{Y}_{n \times k}$ be a matrix containing k response variables, associated with each one of the n samples (the approaches in this work consider only one response variable, therefore $\mathbf{Y}_{n \times 1}$). PLS estimates a p -dimensional ($p \ll n$) space by decomposing \mathbf{X} and \mathbf{Y} into

$$\begin{aligned}\mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^T + \mathbf{F},\end{aligned}$$

where $\mathbf{T}_{n \times p}$ and $\mathbf{U}_{n \times p}$ are matrices composed of the latent variables, matrices $\mathbf{P}_{d \times p}$ and $\mathbf{Q}_{k \times p}$ represent the loadings, and matrices $\mathbf{E}_{n \times d}$ and $\mathbf{F}_{n \times k}$ are the residuals.

The non-linear iterative partial least squares (NIPALS) [1] is the method usually employed for iteratively performing the decomposition defined above by finding projection vectors (also called weight vectors) \mathbf{w}_i and \mathbf{c}_i such that

$$[\text{cov}(\mathbf{t}_i, \mathbf{u}_i)]^2 = \arg \max_{\mathbf{w}_i, \mathbf{c}_i, |\mathbf{w}_i|=|\mathbf{c}_i|=1} [\text{cov}(\mathbf{X}\mathbf{w}_i, \mathbf{Y}\mathbf{c}_i)]^2 \quad (1)$$

where $cov(\mathbf{t}_i, \mathbf{u}_i)$ denotes the sample covariance between vectors \mathbf{t}_i and \mathbf{u}_i [41].

For applying NIPALS, the matrices \mathbf{X} and \mathbf{Y} are first normalized, in which their variables will present zero-mean and one-standard deviation, then they are decomposed by subtracting their rank-one approximations as

$$\begin{aligned}\mathbf{X}_{i+1} &= \mathbf{X}_i - \mathbf{t}_i \mathbf{p}_i^T \\ \mathbf{Y}_{i+1} &= \mathbf{Y}_i - \mathbf{t}_i \mathbf{q}_i^T,\end{aligned}$$

where \mathbf{X}_i and \mathbf{Y}_i are the data representation for the i -th iteration, where $\mathbf{t}_i = \mathbf{X}_i \mathbf{w}_i$ represents the i -th column of matrices \mathbf{T} (\mathbf{w}_i represents the projection vector associated with the i -th latent variable), $\mathbf{X}_1 = \mathbf{X}$ and $\mathbf{Y}_1 = \mathbf{Y}$, and \mathbf{p}_i and \mathbf{q}_i denote the i -th column of the matrices \mathbf{P} and \mathbf{Q} , respectively.

After the extraction of p projection vectors (resulting in the matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p]$), the p -dimensional representation of $\mathbf{X}_{n \times d}$ is given by $\mathbf{T}_{n \times p}$, which is used to extract the regression coefficients $\boldsymbol{\beta}_{d \times k}$ by

$$\boldsymbol{\beta} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{T}^T \mathbf{Y}.$$

Finally, the regression responses, Y_v , for a feature vector $\mathbf{v}_{d \times 1}$ is obtained by

$$Y_v = \bar{Y} + \boldsymbol{\beta}^T \mathbf{v} S,$$

where $\bar{Y}_{1 \times k}$ is the sample mean of each variable of \mathbf{Y} and $S_{1 \times k}$ is the standard deviation of the variables in \mathbf{Y} .

3.2. Feature Extraction

The first stage of the recognition process is to extract feature descriptors from the face region. After cropping and scaling the face to a given size, each

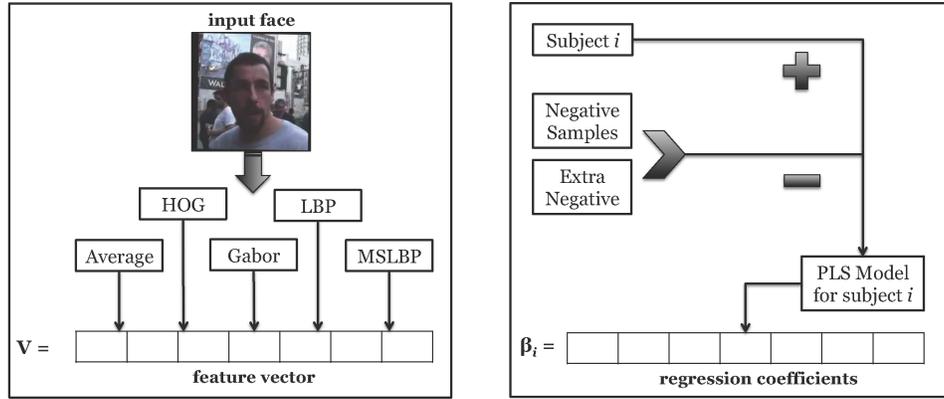
sample is decomposed into a set of overlapping blocks from which feature descriptors are extracted to capture information from multiple types.

We consider a rich combination of different descriptors, including shape information captured by histograms of oriented gradients (HOG) [31], color information captured by the average of pixel colors, salient visual properties extracted using Gabor filters [32] and textural information using the local binary patterns (LBP) descriptor [42], and one of its extensions, the multi-scale local binary patterns (MSLBP) [43]. After the feature extraction for all blocks, the descriptors are concatenated into a feature vector \mathbf{v} , used to describe the face, as illustrated in Figure 2(a).

3.3. Subject Modeling

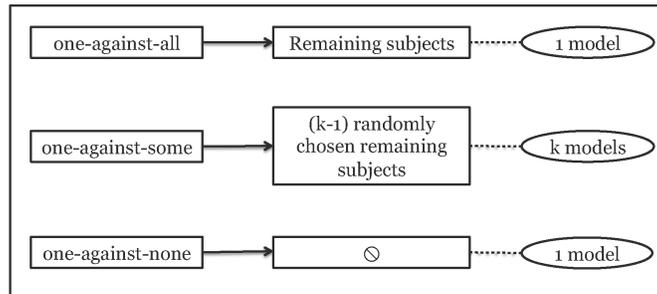
Once the feature extraction is concluded and the face samples are represented in a high dimensional feature space, each subject \mathbf{s}_i in the gallery $g = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\}$, in which \mathbf{s}_i denotes samples of the i -th subject represented by their feature vectors, is modeled by using the Partial Least Squares considering one of the three classification schemes: one-against-all, one-against-some and one-against-none. The general modeling approach is illustrated in Figure 2(b).

The main difference among the three approaches is how the remaining subjects (samples that do not belong to the subject being modeled) are considered as counterexamples. In the first, all remaining subjects are considered to increase the discriminability of the models; in the second, a subset of the remaining subjects is added as counterexamples and multiple models are built for each subject; and, in the third, none remaining subjects are considered and the counterexamples are extracted from an independent set of faces (faces that are not in the gallery). Figure 2(c) illustrates the



(a) Feature extraction. A feature vector is composed by concatenating different types of descriptors.

(b) General subject modeling. Extra negative samples set is unique through all models. On the other hand, negative samples are chosen as described in Figure (c).



(c) Negative sample selection. For each method, it is shown how negative samples are chosen and how many models are built for a training subject.

Figure 2: Diagrams illustrating the training process used in all three techniques.

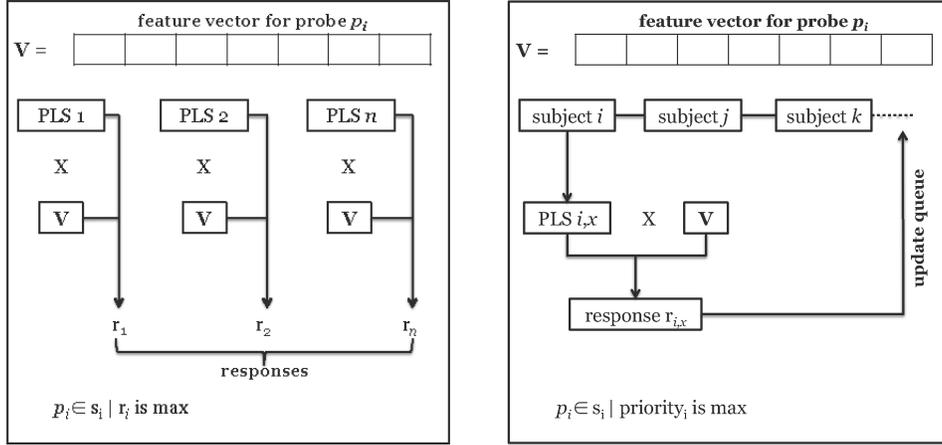
differences.

One-Against-All. To increase the discriminability between classes, when the i -th subject is being modeled, the samples of the remaining subjects, $g \setminus s_i$, are used as counterexamples. The PLS estimates the discrimination ability

of the descriptors in the feature vector and returns regression coefficients β_i . This process is executed for each subject in the gallery. Therefore, at the end, N PLS models will be estimated. The main disadvantage of this approach is that all models need to be rebuilt when a new subject is enrolled.

One-Against-Some. In this approach, not all remaining subjects ($g \setminus \mathbf{s}_i$) are considered as counterexamples, but only $k - 1$ randomly selected remaining subjects are chosen as counterexamples to build the PLS model for the i -th subject. To compensate the reduction in the discriminability, k of such models are learned for an i -th subject. Therefore, it results in $k \times N$ models, where N is the number of known subjects by the time of enrollment and k denotes the number of subjects used as counterexamples.

One-Against-None. As in the previous approaches, the one-against-none creates PLS models for each subject to learn the subject face appearance, resulting in N PLS models. However, instead of adding the remaining subjects as counterexamples, it considers an independent set of faces. Specifically, when a subject is enrolled to the gallery, its PLS model is built by considering its samples as the positive class and only the extra samples as counterexamples. Therefore, while no models need to be rebuilt and the number of models to be stored is N , which are the weakness of the previous approaches, a possible disadvantage of this approach could be the reduced discriminability of the models since no subjects are used as counterexamples. This aspect is evaluated in the experimental results and it turns out to be dependent on the choice of the independent set of faces.



(a) Probe sample matching scheme for one-against-all and one-against-none techniques. The feature vector v of the probe sample is projected against all models resulting in n responses r_i . Then, unknown sample is classified as subject k , where r_k is maximum.

(b) One-against-some probe sample matching scheme. Once priority queue is initialized, feature vector v is projected against model $PLS_{i,x}$, where subject j is the first of the queue and model i,x has not been visited yet. Then, response $r_{i,x}$ is added to the queue if $r_{i,x}$ is lower than currently value. This process is repeated until one of the stop conditions is reached. Finally, the probe sample is classified as the label of the first subject in the queue.

Figure 3: Testing step diagrams. Figure (a) shows classification schemes for one-against-all and one-against-none approaches. Figure (b) describes classification scheme for one-against-some method.

3.4. Probe Sample Identification

After modeling the subjects using the PLS, as described in the previous section, probe samples are identified. When a probe sample is presented to the identification system, its feature vector is used to evaluate the regression response for each PLS model. The subject associated with the highest

regression response (a high regression response indicates that feature vector of the probe and subject’s samples are similar) is considered to be the best match for the probe sample.

Since the one-against-all and one-against-none have only one PLS model per subject, the number of regression evaluations (projections onto the regression coefficient vector) is N , as illustrated in Figure 3(a). However, the one-against-some approach has k models per subject, which leads to $k \times N$ projections, as illustrated in Figure 3(b). To prevent that, a priority queue is employed, described as follows.

Priority Queue for the One-Against-Some. The probe sample matching step aims at finding the best matching using a priority queue to reduce the number of projections needed, since testing all models would cost $k \times N$ projections, opposed as N projections required by the one-against-all and one-against-none approaches.

Aiming at avoiding unnecessary projections (those PLS models associated with subjects unlikely to be the correct match), a priority queue-based approach is proposed. This way, the search relies on testing subjects with high chance of being the correct match. Thus, projections onto models associated with subjects with low priority might be avoided, reducing the computational cost.

The proposed queue associates a priority value to each subject. When a probe sample is presented to the system, the priorities are initialized with the regression response obtained by projecting the probe sample onto a randomly chosen PLS model of each subject (a high response increases the priority). After performing N projections (one per subject), a PLS model for the subject with the highest priority is chosen and the probe is projected

onto it, updating the priority to be the minimum between the current and the regression response obtained. This process is repeated until a stop condition has been reached. Algorithm 1 describes the search process using the priority queue.

Algorithm 1: Priority queue querying

Data: one-against-some models, probe image

Result: classification of probe sample

begin

 initialize empty priority queue;
 project probe to 1 model per subject; // *n* subjects update
 responses into priority queue;

while *unmarked models for top subject* **do**

 select top subject;
 project probe sample to its unmarked model;

if *response is less than current value* **then**

 | update response in the priority queue;

end

 mark current model;

end

 probe is labeled as the highest ranked subject;

end

The use of the priority queue without a stop condition would cause it to test every PLS model. Therefore, a stop condition is incorporated into the system: once all models having the subject with the highest priority as positive class have been considered, the search stops and the probe sample is identified as the subject with the highest priority.

	<i>one-against-all</i>	<i>one-against-some</i>	<i>one-against-none</i>
number of projections	N	$k \times N$ ($\approx N$ with queue)	N
number of PLS models	N	$k \times N$	N
enrollment of a new subject	N	k	1
work without extra samples	yes	yes	no

Table 1: Comparison among the PLS-based classification scheme employed to face identification considering a gallery with N subjects.

Algorithm 1 shows that priorities are defined as the minimum response obtained for each subject. That approach works because regressions responses are positive for the correct class and negative for incorrect classes. Therefore, even the minimum response will be positive for the desired class. This way, after evaluating all models containing a given subject in the positive class, if this subject is still with the highest priority, it is likely to be the best match for the probe sample.

3.5. Discussion

The main drawback of the one-against-all classification scheme is its high computational cost to enroll new subjects, once all existing PLS models need to be rebuilt to consider samples of the new subject as counterexamples. Previous works employing one-against-all approach have accomplished excellent results in terms of accuracy [2, 4] when estimating best match for N known subjects. However, they have failed making the scheme scalable whenever a new subject is presented to the system, as all models require to be rebuilt.

To avoid the reconstruction of all PLS models, we employ two classification schemes: one-against-some and one-against-none. The former builds

PLS models considering only a subset of subjects as counterexamples, which avoids the reconstruction of the PLS models already built, whereas the latter does not consider remaining subjects to build the models.

The intuition to choose the one-against-some approach comes from the fact that while the addition of all remaining subjects as counterexamples might present redundant information to build the PLS model, the application of the pairwise approach may not be enough to emphasize the most discriminative feature descriptors. This way, the one-against-some presents a trade-off between computational cost to add a new subject and high discriminability obtained when all remaining subjects are considered.

The one-against-some approach presents two drawbacks: (1) storage of $k \times N$ models and (2) $k \times N$ projections are required when the priority queue is not employed. Due to that, we propose the classification scheme referred to as *one-against-none* that does not require the creation of extra models. The number of projections per probe sample is N , differently from the one-against-some, and it does not depend on the subjects previously enrolled in the gallery, differently from the one-against-all approach, not requiring the reconstruction of the PLS models.

The one-against-none allows a reduction in the number of models that need to be built and the number of projections when compared to the one-against-some, and the enrollment of new subjects requires the creation of a single model and none has to be rebuilt, differently from the one-against-all and one-against-some approaches. Table 1 summarizes the differences among the approaches. According to the table, the main drawback of the one-against-none approach is that it requires extra samples to allow the creation of the PLS models.

4. Experimental Results

This section evaluates several aspects of the proposed method. First, Section 4.1 describes the data sets used in the experimental validation. Section 4.2 shows results regarding the incremental enrollment of subjects in the gallery. Sections 4.3 and 4.4 evaluate the choice of counterexamples for the one-against-some approach and the generation of counterexamples required by the one-against-none approach, respectively. Finally, Section 4.5 shows the recognition rates achieved in the considered data sets and other results from the literature are shown as reference.

4.1. Data Sets

The proposed method is evaluated on three data sets used for face recognition: FRGC version 1 [44], PubFig83 [45] and YouTube Faces data set [46]¹. These data sets contain different characteristics: FRGC presents a challenging experiment (Experiment 4), which considers few images acquired under controlled conditions and frontal pose for the gallery and images acquired under uncontrolled conditions for the probe, PubFig83 is composed of several uncontrolled images with pose and expression variations, and YouTube Faces contains videos acquired in uncontrolled conditions.

For the FRGC data set, we follow the feature extraction procedure and the evaluation protocols used by Schwartz et al. [2]. The FRGC version 1 for still 2D images considers three experiments, each one with 152 subjects in the gallery: Experiment 1 considers a single controlled sample to build

¹The YouTube Faces data set was originally proposed for face verification, however, we defined a protocol for face identification. The definition of the protocol will be publicly available after this paper acceptance.

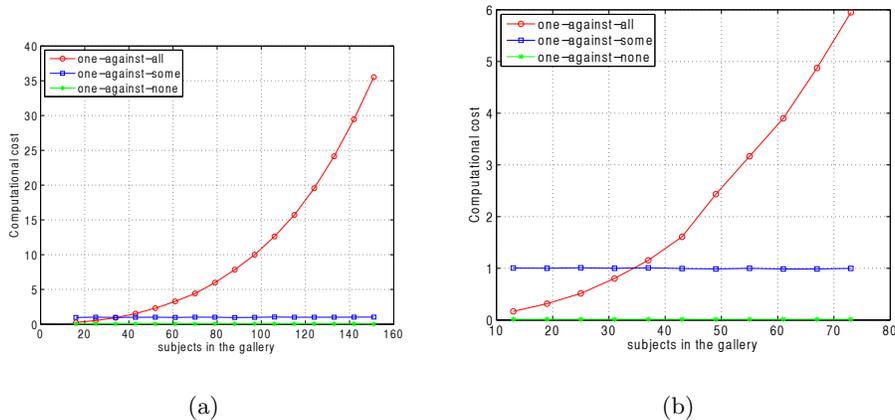


Figure 4: Relative computational cost as a function of the number of subjects in the gallery, enrolled incrementally. (a) enrollment of subjects for the FRGC data set; (b) enrollment of subjects for the PubFig83 data set.

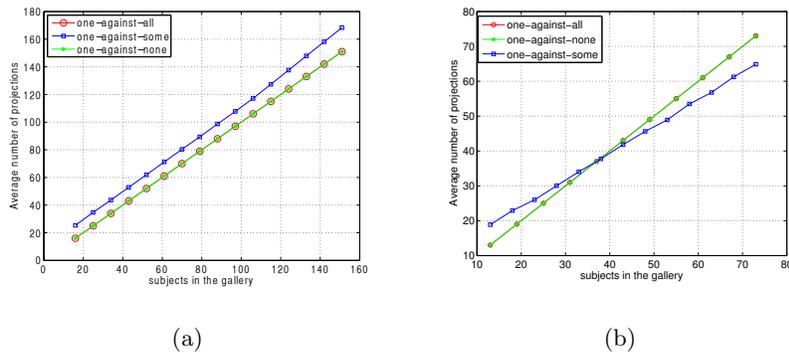


Figure 5: Average number of projections performed to identify a probe sample as a function of the number of subjects in the gallery, enrolled incrementally. (a) average number of projections for the FRGC data set and; (b) average number of projections for the PubFig83 data set.

the gallery and controlled probe images; Experiment 2 considers a gallery with four controlled still images per subjects; and Experiment 4 considers a gallery with a single controlled sample per subject to build the gallery and multiple uncontrolled probe images.

For the PubFig83 data set, composed of 83 different subjects, we follow a slightly different evaluation protocol from that defined by Pinto et al. [45]. In this work, we consider 90 samples per subject for the training and set aside 8 randomly chosen subjects to be used as counterexamples. Due to the large number of samples used in the gallery, we employed the feature extraction procedure defined in [47], which considers fewer descriptors per sample compared to the work in [2]. The samples for the PubFig83 data set were rescaled to 100×100 pixels and the number of descriptors per sample is 6,039. We executed the method ten times, each time with a different split among the samples used to build the gallery, probe samples and extra counterexamples. The average rank-1 recognition rates are then reported.

For the YouTube Faces data set, we propose a new protocol for face identification based on the protocol for face verification defined by Wolf et al. [46]. First of all, each video sequence is downsampled to 8 frames equally distributed, such that there is no bias towards longer or shorter video segments. Then, the database is divided into three subsets: training, testing and extra counterexamples. The training set is composed of one video for each subject that has three videos or more in the data set, resulting in a total number of 533 subjects. The testing set is formed by the same subjects as the training set, though different videos are used (1 or more per subject), resulting in a total of 1,830 videos and 14,640 frames. In addition, the extra counterexamples set is formed by frames of subjects with a single video (without any intersection with the subjects in the training and testing sets). Some frames have been discarded from the counterexamples set to reduce memory consumption once this set is considered to build every model, resulting in 1,123 frames for this set. Finally, samples of the YouTube Faces have been rescaled to 100×100 pixels, resulting in 43,636 descriptors per

sample.

4.2. Incremental Enrollment

An important aspect is the scalability of the enrollment of new subjects in the gallery. Thus, we evaluate the proposed approach focusing on this aspect and compare the results to the original one-against-all approach. For the evaluation, as performed by Carlos et al. [8], we build an initial gallery with a few subjects (in this experiment, 15 for FRGC and 13 for PubFig data) and probe samples are projected onto these models. After the initial enrollment, new subjects are incrementally added to the gallery until all subjects have been added.

The first experiment analyzes the computational cost to enroll the n -th subject in a gallery, considering that $n - 1$ subjects have already been enrolled. The results shown in Figure 4 demonstrate that the enrollment in both one-against-some and one-against-none methods present constant computational time and one-against-all approach takes a quadratic computational time as a function of the number of subjects already enrolled in the gallery. Therefore, one-against-some and one-against-none approaches can be scalable in a dynamic environment, where new subjects are added on-the-fly since the computational cost is invariant to the number of subjects in the gallery. In addition, the enrollment based on the one-against-none demonstrated to be faster than one-against-some by a constant factor.

The second experiment evaluates the average number of projections needed to classify a probe sample as new subjects are enrolled in the gallery. The results presented in Figure 5 show that one-against-all and one-against-none, due to their inherent resemblance in the way models are built, have the same number of projections. On the other hand, the one-against-some

depends on the priority queue to achieve a similar behavior, otherwise the number of projections would be $n \times k$ [8].

Based on the results, we can conclude that the approach proposed in this work is scalable in a dynamic environment where subjects are incrementally added to the gallery and they present a constant time to enroll a new subject and linear time to test a probe sample. A comparison regarding recognition rates will be conducted in Section 4.5.

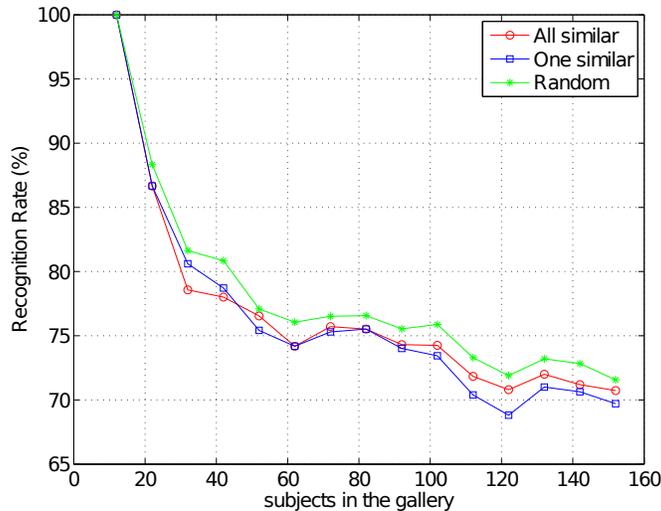


Figure 6: Incremental gallery using one-against-some method for FRGC data set comparing three approaches for selecting the counterexamples when a new individual is presented to the gallery. Considering $k = 10$ [8].

4.3. Similarity Applied to One-Against-Some

The selection of counterexamples for the one-against-some approach is performed by choosing some random subjects currently present in the gallery. Here, we evaluate whether another heuristic to choose the counterexamples would be able to improve the results. Two heuristics have been

considered. The first, referred to as *one similar*, uses as counterexamples the most similar sample to the subject that will be enrolled and choose the $k - 1$ remaining counterexamples randomly. The second heuristic, referred to as *all similar*, uses the k most similar samples to the subject that will be enrolled in the gallery. Figure 6 compares the original approach (random) to these two heuristics for the FRGC data set.

The results in Figure 6 show that the random method still provides higher accuracy, such that it is the approach applied throughout this work. Furthermore, the method using all similar subjects demonstrated to be superior than considering just the most similar, thereby mixing the two approaches (all random and all similar) is not satisfactory.

4.4. Creation of Extra Negative Samples

The one-against-none has proven to be a promising classification scheme. However, to be a viable method, extra counterexamples need to be used. Since not all data set contain samples that can be used as counterexamples, we tested different approaches to generate samples without human supervision, evaluating them with the one-against-none itself and training and testing images from FRGC data set.

We have evaluated four strategies to generate counterexamples and compare them to the results achieved using samples from the training partition of the FRGC data set. First, we collected faces from Google Images website, resulting in a 161 image data set, each one manually cropped to 138×160 pixels to match FRGC image dimensions. In the second strategy, counterexamples were generated by the permutation variables of features between training samples. The next one used a normal distribution to generate feature descriptors for counterexamples, independent on other parameters. The

last approach applied a method for creating counterexamples by using the same mean and standard deviation of each feature of the training samples.

The results achieved by the counterexample generation strategies are shown in Figure 7, where the differences among them are remarkable. The first group is composed of three methods based on randomly selected faces, where the results are almost negligible, as the best result is approximately 10% accuracy. However, the method based on choosing random faces from Google Images demonstrated to be reasonable when compared to the first group, reaching approximately 50% of recognition. The original set, however, is the one with higher rankings. This way, we conclude that the construction of an extra counterexample set is highly dependent on the training/testing set, such that it is difficult to create one randomly. Therefore, the one-against-none method is affected by the construction of an extra counterexample set for the scenario.

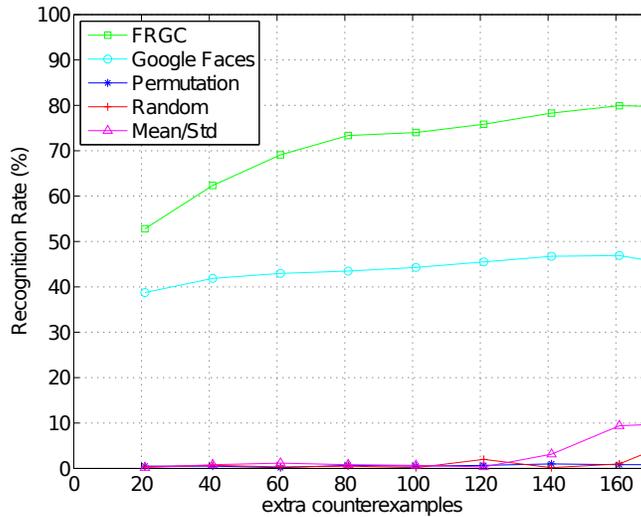


Figure 7: Comparison among methods for creating extra counterexamples by using the one-against-none approach.

4.5. Accuracy Comparison

In Section 4.2, we analyzed computational cost for incremental environments, however, recognition rates are also important since a scalable approach must handle properly increases in the gallery and still be able to classify probe samples correctly. Thus, in this section we compare our method with some state-of-the-art works and some known baselines available in the literature for all three data sets described in Section 4.1.

Results in Tables 2, 3 and 4 show that the proposed method achieves rates very similar to those obtained with one-against-some and one-against-all. However, compared to state-of-the-art methods, the proposed technique produces lower recognition rates since, as explained in Carlos et al. [8], it is limited to the used feature descriptors and the PLS approach defined by the one-against-all method in [2]. For example, the method proposed by Choi et al. [4] considers an enhanced set of feature descriptors and the work proposed by Chiachia et al. [6] considers an SVM classifier associated with biologically inspired feature descriptors.

We also compare our methods with SVM models instead of PLS using the same classification schemes – one-against-all and one-against-none – and the feature descriptors. The results are based on the libSVM implementation [48]. The results show that the proposed one-against-none method is still advantageous when applied to other classification methods since it compensates the small downside of decreased accuracy with the upside of performance gains, as explained in previous sections. In addition, the can see that the results obtained with PLS are significantly better than the results achieved with SVM, even though the same classification schemes and features have been used.

Furthermore, all three PLS-based methods fail to achieve high recogni-

tion rates for YouTube Faces. Since this data set has not been used for face identification, there are no results available for the same protocol we are employing. Therefore, we evaluated the results using the traditional eigenfaces approach [49] and the rank-1 recognition rate are even lower, as shown in Table 4. Such poor results are explained by the complexity of the data set since the faces are poorly aligned and we are employing a holistic feature extraction approach.

Section 4.2 demonstrated that our proposed method is scalable in terms of adding new subjects in the gallery at constant cost and the average number of projections required to test a probe sample is equal to the one-against-all. In this section, we have shown that recognition rate of the one-against-none approach is close to that obtained with one-against-all and one-against-some. Thus, the proposed method is robust and more scalable to be used as an alternative to the original one-against-all approach, which is promising for employing face identification on dynamic scenarios.

5. Conclusions

Most available approaches related to face identification present scalability problems when the number of new subjects in the gallery substantially increases, demanding the reconstruction of the models representing the gallery individuals. Therefore, this work investigates classification schemes to reduce the computation cost when new subjects are enrolled in the gallery with low impact in the accuracy.

A new classification scheme based on PLS models, called one-against-none, was proposed and evaluated in three well-known data sets. A set of face samples that do not belong to subjects in the gallery can be used to build

Table 2: Accuracy comparison (rank-1 recognition rates in %) for the FRGC data set.

Method	Exp.1	Exp.2	Exp.4
SVM multi-class	-	-	21.21
SVM one-against-all	-	-	38.82
SVM one-against-none	-	-	43.91
LC ₁ C ₂ [50]	-	-	75.00
ROCA [51]	-	96.40	75.50
CS-POP [4]	98.00	99.80	89.00
PLS one-against-all [2]	97.90	99.80	86.20
PLS one-against-some [8]	97.20	99.30	84.70
proposed PLS one-against-none	97.20	99.10	83.72

Table 3: Accuracy comparison (rank-1 recognition rates \pm std. err. (%)) for the PubFig83 data set.

Method	Recognition Rate
SVM one-against-all	41.96 \pm 2.35
SVM one-against-none	33.67 \pm 3.28
HT-L3-1st [45]	87.11 \pm 0.56
PS-PLS [6]	88.75 \pm 0.26
PLS one-against-all [2]	67.68 \pm 1.85
PLS one-against-some [8]	66.83 \pm 1.08
proposed PLS one-against-none	66.71 \pm 1.60

the models. Since the scheme allows incremental maintenance of the gallery, face recognition can be performed in an effective and a scalable manner. The proposed approach was compared to two other classification strategies, the

Table 4: Accuracy comparison (rank-1 recognition rates (%)) for the YouTube Faces data set.

Method	Recognition Rate
Eigenfaces	0.50
PLS one-against-all [2]	18.39
PLS one-against-some [8]	17.75
proposed PLS one-against-none	17.51

one-against-all and the one-against-some. Although only a single model has to be built when a new subject is added to the gallery, without any reconstruction of other existing models, the one-against-none scheme presented similar recognition rates compared to the other two schemes, proving to be a promising approach to performing face identification on dynamic environments.

The main difficulty found in the one-against-none approach was the generation of extra samples, used to build the models. To generate such samples, experiments were performed based on different strategies, however, none demonstrated to be effective – the best results were achieved using samples provided by the own data set. Therefore, this point needs to be further investigated.

Acknowledgements

The authors are grateful to FAPESP, FAPEMIG, CNPq for the financial support.

References

- [1] H. Wold, Partial Least Squares, in: Encyclopedia of Statistical Sciences, Vol. 6, Wiley, New York, NY, USA, 1985, pp. 581–591.
- [2] W. R. Schwartz, H. Guo, J. Choi, L. S. Davis, Face Identification Using Large Feature Sets, *IEEE Trans. on Image Processing* 21 (4) (2012) 2245–2255.
- [3] H. Guo, W. R. Schwartz, L. S. Davis, Face Verification using Large Feature Sets and One Shot Similarity, in: International Joint Conference on Biometrics, 2011.
- [4] J. Choi, H. Guo, W. R. Schwartz, L. S. Davis, A Complementary Local Feature Descriptor for Face Identification, in: IEEE Workshop on Applications of Computer Vision, Breckenridge, CO, USA, 2012, pp. 121–128.
- [5] A. Sharma, D. Jacobs, Bypassing Synthesis: PLS for Face Recognition with Pose, Low-Resolution and Sketch, in: IEEE Intl. Conference on Computer Vision and Pattern Recognition, 2011, pp. 593–600.
- [6] G. Chiachia, N. Pinto, W. R. Schwartz, A. Rocha, A. X. Falcao, D. Cox, Person-Specific Subspace Analysis for Unconstrained Familiar Face Identification, in: British Machine Vision Conference, 2012.
- [7] A. Li, S. Shan, X. Chen, W. Gao, Cross-Pose Face Recognition based on Partial Least Squares, *Pattern Recognition Letters* 32 (15) (2011) 1948–1955.
- [8] G. P. Carlos, H. Pedrini, W. R. Schwartz, Fast and Scalable Enrollment for Face Identification based on Partial Least Squares, in: IEEE

International Conference on Automatic Face and Gesture Recognition, 2013.

- [9] X. Zhang, Y. Gao, Face Recognition Across Pose: A Review, *Pattern Recognition* 42 (11) (2009) 2876–2896.
- [10] A. F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D Face Recognition: A Survey, *Pattern Recognition Letters* 28 (14) (2007) 1885–1906.
- [11] Z. Cao, Q. Yin, X. Tang, J. Sun, Face Recognition with Learning-Based Descriptor, in: *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2707–2714.
- [12] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar, Attribute and Simile Classifiers for Face Verification, in: *IEEE Intl. Conference on Computer Vision*, 2009.
- [13] J. Y. Choi, Y. M. Ro, K. N. Plataniotis, Color local texture features for color face recognition, *Image Processing, IEEE Transactions on* 21 (3) (2012) 1366–1380.
- [14] O. Déniz, G. Bueno, J. Salido, F. De la Torre, Face recognition using histograms of oriented gradients, *Pattern Recognition Letters* 32 (12) (2011) 1598–1603.
- [15] N.-S. Vu, Exploring patterns of gradient orientations and magnitudes for face recognition, *Information Forensics and Security, IEEE Transactions on* 8 (2) (2013) 295–304.
- [16] S.-J. Wang, J. Yang, N. Zhang, C.-G. Zhou, Tensor discriminant color

- space for face recognition, *Image Processing, IEEE Transactions on* 20 (9) (2011) 2490–2501.
- [17] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, DeepFace: Closing the Gap to Human-Level Performance in Face Verification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [18] Y. Gao, Y. Qi, Robust Visual Similarity Retrieval in Single Model Face Databases, *Pattern Recognition* 38 (7) (2005) 1009–1020.
- [19] Q. Yuan, A. Thangali, S. Sclaroff, Face Identification by a Cascade of Rejection Classifiers, in: *IEEE Computer Vision and Pattern Recognition Workshops, 2005*, pp. 152–159.
- [20] Z. Zeng, T. Fang, S. Shah, I. Kakadiaris, Local Feature Hashing for Face Recognition, in: *IEEE International Conference on Biometrics: Theory, Applications and Systems, 2009*.
- [21] Q. Cao, Y. Ying, P. Li, Similarity Metric Learning for Face Recognition, in: *IEEE Intl. Conference on Computer Vision, Sydney, Australia, 2013*, pp. 2408–2415.
- [22] Y. Yan, H. Wang, C. Li, C. Yang, B. Zhong, An Effective Unconstrained Correlation Filter and its Kernelization for Face Recognition, *Elsevier Neurocomputing* 119 (2013) 201–211.
- [23] Y. Yan, H. Wang, D. Suter, Multi-Subregion Based Correlation Filter Bank for Robust Face Recognition, *Pattern Recognition* 47 (11) (2014) 3487–3501.
- [24] M. Turk, A. Pentland, Eigenfaces for Recognition, *J. Cognitive Neuroscience* 3 (1991) 71–86.

- [25] K. Etemad, R. Chellappa, Discriminant Analysis for Recognition of Human Face Images, *Journal of Optical Society of America A* 14 (8) (1997) 1724–1733.
- [26] M. Vasilescu, D. Terzopoulos, Multilinear Image Analysis for Facial Recognition, in: *IAPR International Conference on Pattern Recognition*, 2002, pp. 511–514.
- [27] B. Moghaddam, Bayesian Face Recognition, *Pattern Recognition* 13 (11) (2000) 1771–1782.
- [28] C. A. Perez, L. A. Cament, L. E. Castillo, Methodological improvement on local gabor face recognition based on feature selection and enhanced borda count, *Pattern Recognition* 44 (4) (2011) 951–963.
- [29] S. Xie, S. Shan, X. Chen, J. Chen, Fusing local patterns of gabor magnitude and phase for face recognition, *Image Processing, IEEE Transactions on* 19 (5) (2010) 1349–1361.
- [30] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 41 (6) (2011) 765–781.
- [31] N. Dalal, B. Triggs, Histograms of Oriented Gradients for Human Detection, in: *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [32] J. G. Daugman, Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual

Cortical Filters, *Journal of the Optical Society of America A* 2 (1985) 1160–1169.

- [33] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of Gabor Phase Patterns (HGPP): A Novel Object Representation Approach for Face Recognition, *IEEE Trans. on Image Processing* 16 (2007) 57–68.
- [34] Y.-N. Zhang, Y. Xia, Z.-G. Lin, D. D. Feng, 2D Representation of Facial Surfaces for Multi-Pose 3D Face Recognition, *Pattern Recognition Letters* 33.
- [35] B. Yao, H. Ai, S. Lao, Person-Specific Face Recognition in Unconstrained Environments: A Combination of Offline and Online Learning, in: *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1–8.
- [36] H. Zhou, A. H. Sadka, Combining Perceptual Features With Diffusion Distance for Face Recognition, *IEEE Intl. Conference on Systems, Man and Cybernetics* 41 (5) (2011) 577–588.
- [37] R. Rifkin, A. Klautau, In Defence of One-vs-All Classification, *Journal of Machine Learning Research* 5 (2004) 101–141.
- [38] M. Krzysko, W. Wolynski, New Variants of Pairwise Classification, *European Journal of Operational Research* 199 (2) (2009) 512–519.
- [39] S.-H. Park, J. Fürnkranz, Efficient Pairwise Classification, in: *European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, 2007, pp. 658–665.

- [40] W. R. Schwartz, Scalable People Re-Identification Based on a One-Against-Some Classification Scheme, in: IEEE Intl. Conference on Image Processing, 2012.
- [41] R. Rosipal, N. Kramer, Overview and Recent Advances in Partial Least Squares, Lecture Notes in Computer Science 3940 (2006) 34–51.
- [42] T. Ojala, M. Pietikäinen, D. Harwood, A Comparative Study of Texture Measures with Classification based on Featured Distributions, Pattern Recognition 29 (1) (1996) 51–59.
- [43] T. Ojala, M. Pietikäinen, T. Maenpää, Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.
- [44] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. C., K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge, in: IEEE Intl. Conference on Computer Vision and Pattern Recognition, 2005, pp. 947–954.
- [45] N. Pinto, Z. Stone, T. Zickler, D. D. Cox, Scaling Up Biologically-Inspired Computer Vision: A Case Study in Unconstrained Face Recognition on Facebook, in: IEEE Intl. Conference on Computer Vision and Pattern Recognition, 2011.
- [46] L. Wolf, T. Hassner, I. Maoz, Face Recognition in Unconstrained Videos with Matched Background Similarity, in: IEEE Intl. Conference on Computer Vision and Pattern Recognition, 2011, pp. 529–534.

- [47] W. R. Schwartz, H. Guo, L. S. Davis, A Robust and Scalable Approach to Face Identification, in: European Conference on Computer Vision, Vol. 6316 of Lecture Notes in Computer Science, 2010, pp. 476–489.
- [48] R.-E. Fan, P.-H. Chen, C.-J. Lin, Working set selection using second order information for training support vector machines, *The Journal of Machine Learning Research* 6 (2005) 1889–1918.
- [49] L. Sirovich, M. Kirby, Low-Dimensional Procedure for the Characterization of Human Faces, *Journal of the Optical Society of America* 4 (3) (1987) 519–524.
- [50] P. Shih, C. Liu, Evolving Effective Color Features for Improving FRGC Baseline Performance, in: *IEEE Computer Vision and Pattern Recognition Workshops*, 2005, pp. 156–163.
- [51] F. D. la Torre, R. Gross, S. Baker, V. Kumar, Representational Oriented Component Analysis (ROCA) for Face Recognition with One Sample Image per Training Class, in: *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2005, pp. 266 – 273.