Person-Specific Subspace Analysis for Unconstrained Familiar Face Identification

Giovani Chiachia[1,2]
giovanichiachia@gmail.com
Nicolas Pinto[2,3]
pinto@mit.edu
William Robson Schwartz[4]
william@dcc.ufmg.br
Anderson Rocha[1]
anderson.rocha@ic.unicamp.br
Alexandre X. Falcão[1]
afalcao@ic.unicamp.br
David Cox[2]
davidcox@fas.harvard.edu

[1] Institute of Computing
University of Campinas
Campinas, Brazil

[2] Rowland Institute
Harvard University
Cambridge, USA

[3] McGovern Institute
Massachusetts Institute of Technology
Cambridge, USA

[4] Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

Abstract

While significant strides have been made in the recognition of faces under controlled viewing conditions, face recognition "in the wild" remains a challenging unsolved problem [12, 13, 21]. Interestingly, while humans are generally excellent at identifying familiar individuals under such conditions, their performance is significantly worse with unfamiliar individuals [5] and groups [19], leading to the idea that brain may have enhanced or specialized representations of familiar individuals [6]. Inspired by these observations, we explored the use of a number of subspace analysis techniques, applied to various visual representations, to generate person-specific subspaces of "familiar" individuals for face identification. In particular, we introduce a person-specific application of partial least squares (PS-PLS) to generate per-individual subspaces, and show that operating in these subspaces yields state-of-the-art performance on the challenging *PubFig83* familiar face identification benchmark. The results underscore the potential importance of incorporating a notion of familiarity into face recognition systems.

1 Introduction

Over the past two decades, dramatic advances have been made in the performance of face recognition algorithms operating on images acquired under relatively controlled conditions [9]. Indeed, under such conditions, automated face recognition can even surpass human performance in the task of matching pairs of unfamiliar faces [9]. However, in settings where images are less controlled (with variation in view, lighting, expression, *etc*.) and where human subjects are familiar with the faces that are tested (e.g. the faces of celebrities), the advantage of humans over machines is still substantial [15]. While the issue of uncontrolled

© 2011. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

variation has received increased attention in recent years with the creation of the "Labeled Faces in the Wild" dataset [13], the notion of "familiarity" in automated face recognition has been relatively unexplored.

We argue that study of familiarity in computer vision is interesting and timely for at least two reasons. First, the automated recognition of "familiar" faces is increasingly relevant in an age where social media has made available an ever-growing torrent of images of friends and family members. While many current face recognition benchmark sets are organized around deciding whether two probe faces are the same or different, in social media contexts, the problem is often instead that of recognizing which individual – from a fixed gallery of friends, each with many labeled examples – a given face image belongs to. This *identification* problem has a natural relationship to the notion of "familiarity" in human face recognition, in that a large number of past examples of a relatively small cohort of individuals are leveraged to recognize new examples.

Second, a longstanding and growing body of neuroscience and psychology research suggests that human face recognition with familiar and unfamiliar faces is substantially different, possibly even relying on qualitatively different internal representations [5, 6, 19]. Indeed, while human performance with unfamiliar faces is generally poor, performance with familiar faces is excellent, and is superior even in tasks that do not depend on the identity of the faces being observed (*e.g.* determining whether a face is male or female) [2, 25]. Understanding how past experience with a particular set of faces can be leveraged to yield better recognition performance in unconstrained environments has the potential to greatly improve our understanding of the fundamental mechanisms and constraints of face recognition.

Inspired by the idea that humans may rely on enhanced representations for familiar individuals, we here explore the construction of person-specific subspaces for performing face identification. The creation of subspaces tailored for faces is a classic technique in the face recognition literature; a variety of matrix-factorization techniques have been applied to faces (*e.g. Eigenfaces* [31], *Fisherfaces* [3], *Tensorfaces* [33], *etc.*), which seek to model structure across a set of training faces, such that new face examples can be projected into these spaces and can be compared. A principle advantage of projecting into such subspaces is in the reduction of noise by limiting comparison to few relevant dimensions of variability in faces, as measured across a large number of images. However, while these methods naturally capture general structure across a set of faces, they typically discover either just structure that is common to all faces (as in the case of eigenfaces), or just structure that is discriminative between two sets of faces (as in the case of Fisherfaces). Furthermore, neither of these classic methods have been shown to perform particularly well (by themselves at least) in the case of unconstrained face images.[1]

One attempt to model familiar faces with computational systems is presented in [7], where the authors considered unconstrained face pictures of celebrities taken on different occasions. Face identity was modeled by averaging up to 19 carefully registered images of each celebrity and then applying principal components analysis (PCA) in the average samples. In their experimental setup, the authors showed that averaging face images of the individuals leads to better identification performance than considering their images separately. Taking this idea further, Burton *et al.* [6] argued that despite the fact that face averaging seems to eliminate information irrelevant to identity, "true" variance in appearance is possibly informative, and should be embedded in the matching process rather than eliminated. They also suggest that the mental representation of familiar faces seems to be specific to each individ-

[1] http://vis-www.cs.umass.edu/lfw/results.html

ual. To test this idea, they performed PCA on a set of 48 carefully registered unconstrained face images of *one* celebrity, and provided insights in this direction.

Here, we build person-specific face subspaces from orthonormal projection vectors obtained using a per-individual configuration of partial least squares [34], which we refer to as "person-specific" PLS or *PS-PLS* models. A key motivating idea for this work is that such person-specific subspaces, due to its supervised nature, can capture both those aspects of the face that are good for discriminating it from others, as well as natural variation in appearance that is present in the unconstrained images of that individual.

While partial least squares methods have been used in other contexts in face recognition before [11, 28], in the absence of a dataset containing many examples per individual, it is not possible for PLS methods to model natural variability in face appearance found in unconstrained images. To overcome this problem, we use the *Pubfig83* dataset, a subset of the *Pubfig* face dataset [15] reconfigured for the problem of unconstrained face identification [22].

In order to evaluate our method, we replicate the previous best results with the *PubFig83* data set [22] and consider them as baselines. The baseline methods consist of binary linear support vector machines (SVMs) trained on different visual representations of faces in a one-versus-all multi-class setting. To compare these methods, we project feature descriptor vectors from this method into custom PS-PLS subspaces that we construct so that each binary linear SVM is trained in a different and person-specific space corresponding to the positive class.

Within this framework, we also compare our approach with subspaces built via other linear techniques. In particular, we consider person-specific principal component analysis (PCA), similar in spirit to the approach of Burton *et al.* [6], along with traditional non-person-specific PCA and linear discriminant analysis (LDA), since they are well-known in literature and because PLS shares properties with each. We further consider models obtained by random projections, which have been shown to perform surprisingly well (*e.g.* [35]). Special attention is given to parameters that could bias the results, namely, the number of projection vectors of each model and the SVM regularization parameter C. Finally, as an additional test of these methods, we evaluate the approach on the *Facebook100* dataset [22], which is constructed from a large set of real-world face images taken from the Facebook social network.

2 Partial Least Squares (PLS)

Partial least squares is a class of methods primarily designed to model relations between sets of observed variables by means of latent vectors [23, 34]. It can also be applied as a discriminant tool for the estimation of a low dimensional space that maximizes the separation between samples of different classes. PLS has been used in different areas [16, 18] and, recently, it is also being successfully applied to Computer Vision problems for dimensionality reduction, regression and classification purposes [11, 14, 27, 28, 29, 30].

Given two matrices $\mathbf{X}$ and $\mathbf{Y}$ respectively with d and k mean-centered variables and both with n samples, PLS decomposes $\mathbf{X}$ and $\mathbf{Y}$ into

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad \text{and} \quad \mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}, \tag{1}$$

where $\mathbf{T}_{n\times p}$ and $\mathbf{U}_{n\times p}$ are matrices containing the desired number p of latent vectors, 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.

One approach to perform the PLS decomposition employs the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm [34], in which projection vectors $\mathbf{w}$ and $\mathbf{c}$ are determined iteratively such that

$$[cov(\mathbf{t},\mathbf{u})]^2 = \max_{||\mathbf{w}||=||\mathbf{c}||=1} [cov(\mathbf{Xw},\mathbf{Yc})]^2, \tag{2}$$

where $cov(\mathbf{t},\mathbf{u})$ is the sample covariance between the latent vectors $\mathbf{t}$ and $\mathbf{u}$. In order to compute $\mathbf{w}$ and $\mathbf{c}$, given a random initialization of $\mathbf{u}$, the following steps are repeatedly executed [23]:

1) $\mathbf{u}_{old} = \mathbf{u}$
2) $\mathbf{w} = \mathbf{X}^T\mathbf{u}$
3) $||\mathbf{w}|| \to 1$
4) $\mathbf{t} = \mathbf{Xw}$
5) $\mathbf{c} = \mathbf{Y}^T\mathbf{t}$
6) $||\mathbf{c}|| \to 1$
7) $\mathbf{u} = \mathbf{Yc}$
8) if $||\mathbf{u} - \mathbf{u}_{old}|| > \varepsilon$, go to step 1

When there is only one variable in $\mathbf{Y}$, *i.e.*, if $k = 1$, then $\mathbf{u}$ can be initialized as $\mathbf{u} = \mathbf{Y} = \mathbf{y}$. In this case, the steps above are executed only once per latent vector to be extracted [23]. The loadings are then computed by regressing $\mathbf{X}$ on $\mathbf{t}$ and $\mathbf{Y}$ on $\mathbf{u}$, *i.e.*,

$$\mathbf{p} = \mathbf{X}^T\mathbf{t}/(\mathbf{t}^\mathbf{T}\mathbf{t}) \quad \text{and} \quad \mathbf{q} = \mathbf{Y}^T\mathbf{u}/(\mathbf{u}^\mathbf{T}\mathbf{u}). \tag{3}$$

In this work, we use PLS to model the relations between face samples and their identities. The relationship between $\mathbf{X}$ and $\mathbf{Y}$ is then *asymmetric* and the *predicted* variables in $\mathbf{Y}$ are modeled as indicators. In the asymmetric case, after computing the latent vectors, matrices $\mathbf{X}$ and $\mathbf{Y}$ are deflated by subtracting their rank-one approximations based on $\mathbf{t}$, that is,

$$\mathbf{X} = \mathbf{X} - \mathbf{tp}^T \quad \text{and} \quad \mathbf{Y} = \mathbf{Y} - \mathbf{tt}^T\mathbf{Y}/(\mathbf{t}^\mathbf{T}\mathbf{t}). \tag{4}$$

Such deflation rule ensures orthogonality among the latent vectors $\{\mathbf{t}_i\}_{i=1}^p$ extracted over the iterations. For details about the different types of PLS, their applicability to regression and other problems, and how they compare with other techniques, we refer the reader to [1, 10, 23].

3 Familiar Face Representation via Person-Specific PLS

From a machine learning perspective, any technique that attempts to discriminate between face identities can be considered "person-specific" in some sense. However, subspace models can offer more degrees of freedom to accommodate within-class variance in appearance. In order to cope with this idea, we model the faces of each person c at a time by setting $k = 1$, $\mathbf{Y}_{n\times k} = \mathbf{y}_c$, and $y_{cs} = 1$ if sample s (out of n) belongs to class c or $y_{cs} = 0$ otherwise. As $\mathbf{Y}$ has a single variable, this variant of PLS is also known as PLS1 [23]. It is worth recalling from Section 2 that when $k = 1$, we can initialize $\mathbf{u} = \mathbf{y}_c$ and that, in this case, obtaining the projection vectors $\{\mathbf{w}\}_{i=1}^p$ is straightforward. In other words, at each iteration i,

$$\mathbf{w}_i = \mathbf{X}i^T\mathbf{y}_c, \tag{5}$$

where $\mathbf{X}i$ is the matrix $\mathbf{X}$ deflated up to iteration i according to Eq. 4.

The person-specific face model that we consider in this work is the subspace spanned by the set of orthonormal vectors $\{\mathbf{w}_i\}_{i=1}^p$ produced by NIPALS for a person c. Given that the variables in $\mathbf{X}$ are also normalized to unit variance, $\mathbf{w}_i$ expresses the relative importance of the face features (*i.e.*, the variables) to discriminate person c from the others. As $\{\mathbf{w}_i\}_{i=1}^p$ are

orthogonal, this model accounts for within-person variance in the face appearance throughout the samples, a property also suggested to be relevant in mental representations of familiar faces [6].

Having obtained a subspace for each face, identities can be predicted with any classification engine that can operate in each subspace (*e.g.*, one-versus-all SVMs). As the dimensionality of person-specific subspaces is expected to be much lower than of the classical multiclass subspace analysis, subsequent learning algorithms may also tend to achieve better generalization [32].

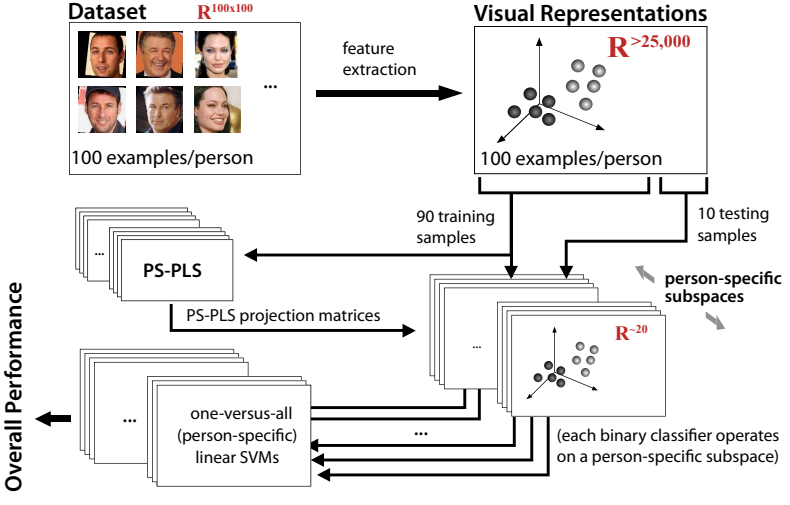

Fig. 1: Our approach. From the training samples, PS-PLS creates a different face subspace for each individual. A different classifier is then trained in each subspace. See text for more details.

In Fig.1 we illustrate our approach. From the visual representation of the training samples, PS-PLS creates a different face subspace for each individual. All training samples are projected onto each person-specific subspace, so that a classifier can be trained by considering the different representations of the samples over the subspaces. The classification engine that we use in this work is made by linear SVMs in a one-versus-all configuration. Given a test sample, an overall decision is made according to decisions made in each person-specific subspace. In this work, we predict the face identity by choosing the person whose corresponding SVM scored highest.

4 Experimental Setup

4.1 Datasets and Visual Representations

The primary dataset that we use in this work is *Pubfig83* [22]. This dataset is a subset of the *Pubfig* dataset [15], which is, in turn, a large collection of real-world images of celebrities collected from the Internet. This subset was established and released to promote research on familiar face recognition from unconstrained images, and it is the result of a series of processing steps aimed at removing spurious face samples from *Pubfig*, *i.e.*, non-OpenCV-detectable, near-duplicate, *etc*. In addition, only persons for whom 100 or more face images remained were considered, leading to a dataset with 83 individuals.

The *Facebook100* dataset is similar in spirit to *Pubfig83*, and a remarkably linear relationship between performance achieved on each set by a variety of algorithms has been reported in [22]. Both sets enable the investigation of face recognition methods where a considerable number of natural face images from the individuals is available. Details about this set can be found in [22]. In both datasets, we are considering the "aligned" version, in which faces images have been pre-aligned by affine-warping [22].

Four different visual representations are considered in the evaluation of our approach. The first three are V1-like+, HT-L2-1st, and HT-L3-1st, which are taken from [22]. These representations are biologically-inspired, and can be thought of as visual models of increasing complexity. The fourth visual representation is a blend of local binary patterns (LBP), histogram of oriented gradients (HOG), and Gabor wavelets (LBP+HOG+Gab), replicated from [28].

4.2 Compared Methods

The main baseline consists of training linear SVMs straight from the visual representation of the images. We call this representation "RAW", and the linear model considered in this case can be though of as an identity matrix. In addition to the performance comparison between RAW and our PS-PLS method, we also consider subspace models obtained via PCA, LDA, and Random Projection (RP). PCA is intuitively appealing in the context of face recognition and decomposes the train set in a way that most of the variance among the samples can be explained by a much smaller and ordered vector basis. LDA is another well-known technique that attempts to separate samples from different classes by means of projection vectors pointing to directions that decrease within-class variance while increasing the between-classes variance. As our PS-PLS setup seeks to maximize the separation only between-class, we argue that this offers a good compromise between LDA and PCA. Finally, due to its interesting properties [4, 35], we also consider RP vectors sampled from a univariate normal distribution.

We further evaluate *person-specific* PCA models (PS-PCA) and multiclass PLS models with the idea that they would provide insight regarding the value of person-specific spaces. PS-PCA models are built only with the train samples of the person. For the multiclass PLS models, we assume k as the number of classes and make $\mathbf{Y}_{n\times k} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$, with $y_{cs} = 1$ if sample s belongs to class c or $y_{cs} = 0$ otherwise. Still, in the inner loop of the NIPALS algorithm, each projection vector is considered after satisfying a convergence tolerance $\varepsilon = 10^{-6}$ or after 30 iterations, whichever comes first (see Section 2 for details). In this case, as $\mathbf{Y}$ has multiple variables, this form of PLS is also known as PLS2 [23].

While there remains substantial room to evaluate other subspace methods (including kernelized versions of PCA [26], LDA [17], and PLS [24]), we chose here to focus on some of the most popular and straightforward methods available, with the goal of cleanly assessing the benefit of building person-specific subspaces to model face familiarity. Future work could compare additional methods.

For all methods, the Scikit-learn package [20] was used to compute the linear models and LIBSVM [8] was used to train the linear SVMs. In all cases, the data was scaled to zero mean and unit variance.

4.3 Evaluation Protocol

We follow the evaluation protocol of [22], reporting the average classification accuracy and standard error for ten random train/test splits of the data. The number of test face images per person is fixed at 10, and here we always consider 90 training examples per individual so that we can evaluate how much we enhance the representation of the faces in the "highest level" of familiarity.

For each dataset, we consider the same splitting rule across the different visual representations. All subspace analysis techniques that we compare can be though of as simple replacements for PS-PLS in Figure 1.

The evaluation framework has two parameters: the regularization constant C of the linear SVMs, and the number of projection vectors to be considered, which is relevant in the cases where the projection vectors are ordered by their variance or discriminative power (PCA, PS-PCA, PLS, and PS-PLS). We use a separate *grid search* to estimate these parameters for each split. For this purpose, we re-split the train set so that we obtain 80 samples per class to generate *intermediate* models and 10 samples per class to *validate* them. We consider $\{10^{-3}, 10^{-2}, \ldots, 10^{6}\}$ as possible values to search for C. For the RAW and LDA models, this is the only parameter that we have to search, because, in the RAW case, no projection is made in practice and, in LDA, the number of projection vectors is fixed to the number of classes minus 1.

The possible number of projection vectors that we consider in the search can be represented as $\{1m, 2m, \ldots, 8m\}$. For person-specific subspace models, $m = 10$, *i.e.*, starting from 10, the number of projection vectors is increased by 10 up to the total number of data points per person in the *validation* set. Correspondingly, for the multiclass models, $m = 10n$, where n is the number of persons in the dataset. The only exception is PLS, where $m = n$. Although PLS is a multiclass model, we observed that the ideal number of projection vectors is concentrated in the first few, and so we decided to refine the search accordingly, while keeping the same number of trials as for the other models.

5 Results

The results obtained on the *Pubig83* dataset are shown in Table 1. In general, comparisons are done with the first row, where performance is assessed with the RAW visual representations [22]. The remaining rows are divided according to the type of subspace analysis technique. It is possible to observe that the only face subspace in which we could consistently get better results than RAW across the different representations is PS-PLS.

With the *multiclass unsupervised* techniques, we see no boost in performance above RAW. Since unconstrained face images have a considerable amount of noise and these techniques do not regard its removal while estimating the models, this is perfectly reasonable. We observe that the drop in performance with RP is related to the fixed maximum number of projection vectors that we consider in grid search (6,640) in proportion to the dimensionality of the input spaces (HT-L2-1st $\approx$ HT-L3-1st $\ll$ V1-like+ $<$ LBP+HOG+Gab). Both for RP and PCA, the most frequently number of projection vectors found by grid search was 6,640. This also gives us the intuition that, operating in these unconstrained face images, the best that PCA can do is to retain as much variance in the input space as possible.

For the *multiclass supervised* subspace models, we observe performance increases with V1-like+ and LBP+HOG+Gab representations, especially for LDA. For HT-L2-1st and HT-

Models	V1-like+	HT-L2-1st	HT-L3-1st	LBP+HOG+Gab	
RAW	74.81 ± 0.35	83.66 ± 0.55	87.66 ± 0.29	82.63 ± 0.28	n ($\mathbb{R}^n$)
Multiclass Unsupervised					
RP	69.04 ± 0.44	79.92 ± 0.50	85.61 ± 0.37	75.07 ± 0.37	6,640
PCA	74.59 ± 0.36	83.36 ± 0.47	87.50 ± 0.28	82.44 ± 0.34	6,640
Multiclass Supervised					
LDA	76.16 ± 0.50	81.14 ± 0.30	85.72 ± 0.33	83.40 ± 0.22	–
PLS	74.90 ± 0.45	83.07 ± 0.47	86.63 ± 0.35	83.02 ± 0.26	332
Person-Specific					
PS-PCA	29.95 ± 0.31	44.76 ± 0.45	52.65 ± 0.62	33.02 ± 0.39	80
PS-PLS	77.59 ± 0.53	84.32 ± 0.38	88.75 ± 0.26	85.42 ± 0.29	20

Table 1: Comparison of different face subspace analysis techniques on the Pubfig83 dataset. In all cases, the final identities are estimated by linear SVMs. In the last column, we present the most frequently number of projection vectors found by grid search (see Section 4.3 for details).

Models	V1-like+	HT-L2-1st	HT-L3-1st
RAW	79.96 ± 0.19	85.81 ± 0.29	88.89 ± 0.25
PCA	79.81 ± 0.18	85.70 ± 0.29	88.88 ± 0.25
LDA	81.04 ± 0.29	83.07 ± 0.26	87.25 ± 0.29
PS-PLS	81.53 ± 0.25	86.84 ± 0.19	89.70 ± 0.25

Table 2: Comparison of different face subspace analysis techniques in the Facebook100 dataset.

L3-1st, there was no improvement, which we suspect may simply result from there being less room for improvement in these cases. When compared with PS-PLS, we think that person-specific manifolds in the multiclass subspace are impaired by a more complex relation among the projection vectors. Since both PLS and PS-PLS follow the same rule for the estimation of the projection vectors, the results corroborate the idea that representing each individual in its own subspace results better performance.

In the *person-specific* category, we see that PS-PCA considerably diminishes the predictive power of the features in the input space. In all cases, the best number of projection vectors found by grid search was 80, *i.e.*, the maximum allowed. When compared with PS-PLS, we can see here the importance of person-specific models being also discriminative, besides generative, for this task. We cannot disregard noise in the unconstrained scenario.

For the *Facebook100* dataset, we present in Table 2 the performance obtained with the most competitive method of each category considered in Table 1. The results are similar to the ones obtained on *Pubfig83*, where PCA representations performed most like RAW, LDA did better in V1-like+, and PS-PLS performed best across all representations.

6 Discussion

In this work, we propose a person-specific application of partial least squares (PS-PLS) to generate per-individual subspaces of familiar faces. By means of a straightforward evaluation methodology, we compared different subspace analysis techniques for modeling the

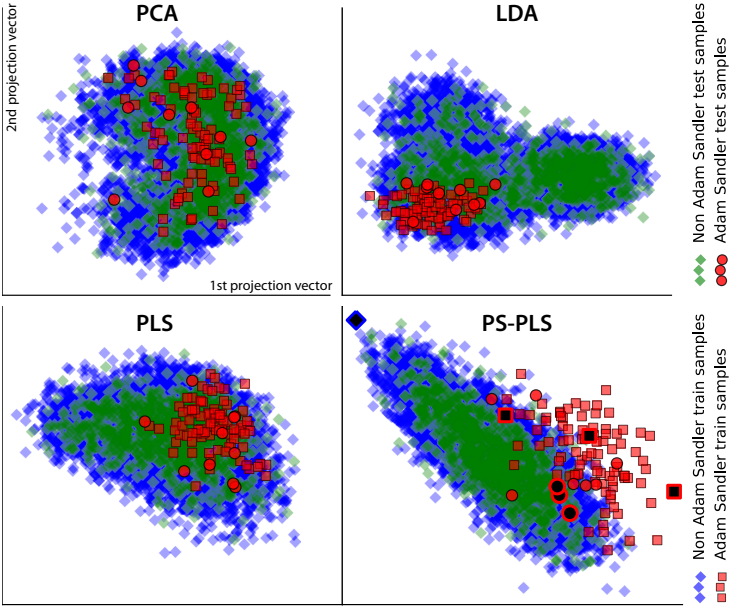

Fig. 2: Visualization of the train and test samples projected onto the first two projection vectors of each model. All models were obtained from the same train/test split using V1-like+. The highlighted samples in PS-PLS are shown in the first row of Fig.3.

problem. We showed that the PS-PLS method consistently performed better than the other subspace models and, here, we seek to understand these models better.

In Fig.2, we present a scatter plot of train and test samples projected onto the first two projection vectors of each model. Adam Sandler's samples are in red[2]. The overall distribution of the points is in accordance to our expectations, where samples are spread out in PCA subspace, are more concentrated, apart with respect to the other classes, and Gaussian shaped in LDA subspace, and are also apart but less concentrated in PLS and PS-PLS.

We note that the predictive power of the first PS-PLS projection vectors is higher than that of the second one. Indeed, in PS-PLS, we found that the only projection vector that leads to mean projection responses significantly different between positive and negative samples is the first one. Although all subsequent projection vectors considerably increase performance, we believe that, from the second vector on, they progressively account more for person-specific variance than discriminative information. In our experiments, performance began to saturate around 20 projection vectors.

Highlighted samples in the PS-PLS scatter plot are shown in the first row of Fig.3, where we also present images of two other individuals. The first column (a) is the result of mapping the importance of each V1-like+ feature back to the spatial domain, based on their relative importance found by the first PS-PLS projection vector. Based on these illustrations, we can roughly see that higher importance is being given to Adam Sandler's mouth and forehead (first row), to Alec Baldwin's eyes, hairstyle, and chin (second row), and to the configural relationship of Angelina Jolie's face attributes (third row).

Columns in (b) of Fig.3 show the person-specific most, average, and least responsive face samples with respect to the projection onto the first PS-PLS projection vector (a). For Adam

[2]As Pubfig83 is a dataset with celebrities, we use their names in the discussion.

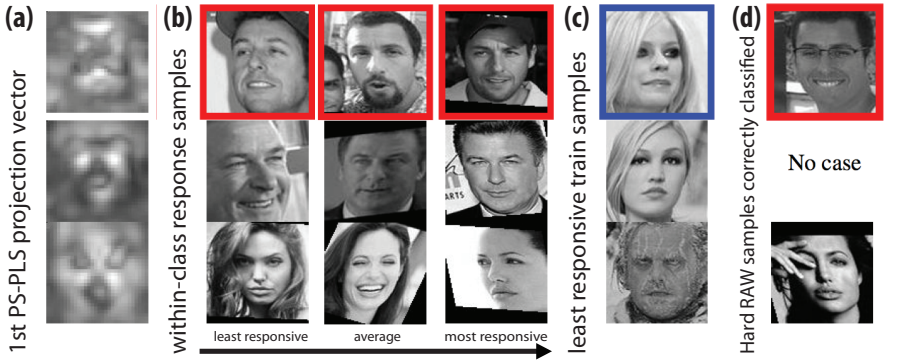

Fig. 3: V1-like+ first PS-PLS projection vectors and representative samples (see Sec.6). The red and blue samples in the first row are the ones highlighted in Fig.2's PS-PLS plot; (b) and (c) refer to the highlighted squares, while (d) illustrates one of the highlighted circles.

Sandler, these samples are highlighted in the PS-PLS scatter plot of Fig.2. It is difficult to infer anything concrete from these images, but we can see that the least responsive samples represent large variations in pose alignment and occlusion.

Still in Fig.3, column (c) represents the overall least responsive training sample with respect to (a). These samples tend to be of the opposite gender, and hair seems to play a role for the first two individuals. Finally, in column (d) we present one test sample of each person that was not recognized when considering the RAW description of the faces, but that was recognized with the aid of PS-PLS models. Despite showing just one sample for Adam Sandler, there were three such cases, which are highlighted in the PS-PLS plot of Fig.2.

In general, we argue that these subspaces are useful both for noise removal and for accentuating discriminative person-specific face aspects (Fig.3a). In the unconstrained familiar face identification setting, both of these issues are of fundamental importance. Considering the results obtained with the RAW visual representations, we see that linear SVMs achieve reasonably high level of performance; however, when these same classifiers are trained and operate in PS-PLS subspaces, they perform better, suggesting that these 20-dimensional person-specific subspaces not only embed comparable levels of the available face identity information, but also amplify it.

7 Acknowledgments

This work was funded in part by FAPESP (10/00994-8, 10/05647-4, 07/52015-0), CNPq (303673/2010-9), Microsoft, the National Science Foundation (IIS 0963668), Google Research, and by The Rowland Institute at Harvard.

References

[1] H. Abdi. Partial Least Squares regression and projection on latent structure regression. *Computational Statistics*, 2010.

[2] B. Balas, D. Cox, and E. Conwell. The effect of real-world personal familiarity on the speed of face information processing. *PloS One*, 2007.

[3] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE PAMI*, 1997.

[4] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. *ACM SIGKDD*, 2001.

[5] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce. Face recognition in poor-quality video: Evidence from security surveillance. *Psych. Science*, 1999.

[6] A. M. Burton, R. Jenkins, and S. R. Schweinberger. Mental representations of familiar faces. *British Journal of Psychology*, 2011.

[7] A.M. Burton, R. Jenkins, P.J.B Hancock, and D. White. Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 2005.

[8] C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2011.

[9] P.J. Phillips et al. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE TPAMI*, 2010.

[10] P. Geladi. Notes on the history and nature of partial least squares (PLS) modelling. *Journal of Chemometrics*, 1988.

[11] H. Guo, W. R. Schwartz, and L. S. Davis. Face verification using large feature sets and one shot similarity. *IEEE IJCB*, 2011.

[12] C. Huang, S. Zhu, and K. Yu. Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. Technical Report TR115, NEC Labs, 2011.

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild. Technical report, Univ. of Massachusetts, Amherst, 2007.

[14] A. Kembhavi, D. Harwood, and L.S. Davis. Vehicle detection using partial least squares. *IEEE PAMI*, 2011.

[15] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. *IEEE ICCV*, 2009.

[16] F. Lindgren, P. Geladi, A. Berglund, M. Sjostrom, and S. Wold. Interactive variable selection (IVS) for PLS. Part II: Chemical Applications. *Journal of Chemometrics*, 1995.

[17] S. Mika, G. Ratsch, J. Weston, B. Schölkopf, and K. R. Mullers. Fisher discriminant analysis with kernels. In *IEEE Neural Networks for Signal Processing Workshop*, pages 41–48, 1999.

[18] D. V. Nguyen and D. M. Rocke. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, 2002.

[19] A. O'toole, K. Deffenbacher, D. Valentin, and H. Abdi. Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, 1994.

[20] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *JMLR*, 2011.

[21] N. Pinto and D. D. Cox. Beyond Simple Features: A Large-Scale Feature Search Approach to Unconstrained Face Recognition. *IEEE FG*, 2011.

[22] N. Pinto, Z. Stone, T. Zickler, and D. D. Cox. Scaling-up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. *IEEE CVPR*, 2011.

[23] R. Rosipal and N. Kramer. Overview and recent advances in partial least squares. *LNCS*, 2006.

[24] R. Rosipal and L. J. Trejo. Kernel partial least squares regression in reproducing kernel hilbert space. *JMLR*, 2:97–123, 2001.

[25] B. Rossion, C. Schiltz, and M. Crommelinck. The functionally defined "face areas" are sensitive to long-term visual familiarity. *Neuroimage*, 2003.

[26] B. Schölkopf, A. J. Smola, and K. R. Müller. Kernel principal component analysis. In *ICANN*, pages 583–588. Springer, 1997.

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

[28] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis. Face identification using large feature sets. *IEEE TIP*, 2011.

[29] W. R. Schwartz, A. Rocha, and H. Pedrini. Face spoofing detection through partial least squares and low-level descriptors. *IEEE IJCB*, 2011.

[30] A. Sharma and D. Jacobs. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. *IEEE CVPR*, 2011.

[31] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. *IEEE CVPR*, 1991.

[32] V.N. Vapnik. An overview of statistical learning theory. *IEEE TNN*, 1999.

[33] M. A. O. Vasilescu and D. Terzopoulos. Multilinear Image Analysis for Facial Recognition. In *IEEE ICPR*, volume 2, pages 511–514, 2002.

[34] H. Wold. Partial least squares. *Encyclopedia of Statistical Sciences*, 1985.

[35] J. Wright and G. Hua. Implicit elastic matching with random projections for pose-variant face recognition. *IEEE CVPR*, 2009.