Robust joint and individual variance explained

Christos Sagonas, Yannis Panagakis, Alina Leidinger, Stefanos Zafeiriou

Angaben zur Veröffentlichung / Publication details:

Sagonas, Christos, Yannis Panagakis, Alina Leidinger, and Stefanos Zafeiriou. 2017. "Robust joint and individual variance explained." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, Honolulu, HI, USA*, edited by Yanxi Liu, James M. Rehg, Camillo J. Taylor, Ying Wu, Rama Chellappa, Anthony Hoogs, and Zhengyou Zhang, 5739–48. Piscataway, NJ: IEEE. https://doi.org/10.1109/cvpr.2017.608.

Robust Joint and Individual Variance Explained

Christos Sagonas[1,2], Yannis Panagakis[1,3], Alina Leidinger[1], Stefanos Zafeiriou[1]
[1]Imperial College London, UK
[2]Onfido, UK
[3] Middlesex University London, UK
christos.sagonas@onfido.com, {i.panagakis, s.zafeiriou}@imperial.ac.uk

Abstract

Discovering the common (joint) and individual subspaces is crucial for analysis of multiple data sets, including multi-view and multi-modal data. Several statistical machine learning methods have been developed for discovering the common features across multiple data sets. The most well studied family of the methods is that of Canonical Correlation Analysis (CCA) and its variants. Even though the CCA is a powerful tool, it has several drawbacks that render its application challenging for computer vision applications. That is, it discovers only common features and not individual ones, and it is sensitive to gross errors present in visual data. Recently, efforts have been made in order to develop methods that discover individual and common components. Nevertheless, these methods are mainly applicable in two sets of data. In this paper, we investigate the use of a recently proposed statistical method, the so-called Joint and Individual Variance Explained (JIVE) method, for the recovery of joint and individual components in an arbitrary number of data sets. Since, the JIVE is not robust to gross errors, we propose alternatives, which are both robust to non-Gaussian noise of large magnitude, as well as able to automatically find the rank of the individual components. We demonstrate the effectiveness of the proposed approach to two computer vision applications, namely facial expression synthesis and face age progression in-the-wild.

1. Introduction

Extracting the modes of variation from two or multiple data sets has created a wealth of research in statistics, signal and image processing, and computer vision. Two mathematically similar but conceptually different models underlie the bulk of the methodologies. In particular, the Canonical Correlation Analysis (CCA) [12] and its variants e.g.,[17, 24] are the methods of choice for extracting linear correlated components among two or more sets of variables. Similarly, inter-battery factor analysis [37] and its extensions e.g., [18] determines the common factors among two sets of variables.

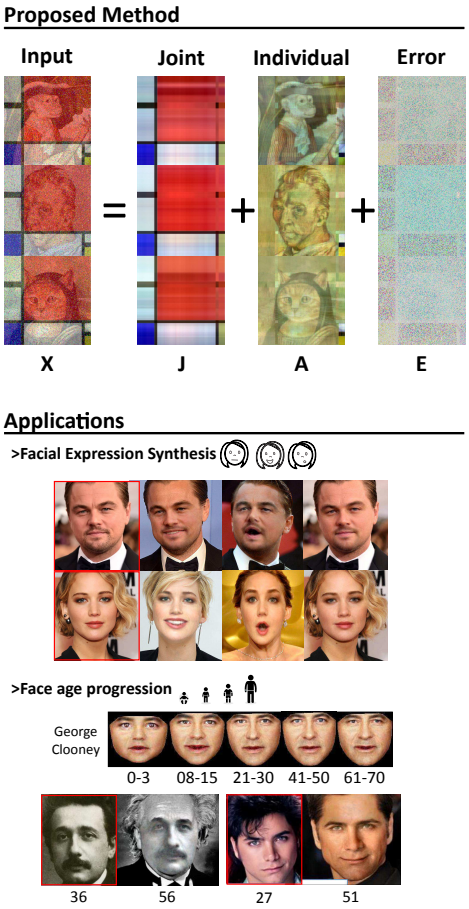


Figure 1. A visual representation of the RJIVE decomposition and its applications considered in this paper. Images highlighted in red boxes are given as input to the RJIVE (best viewed in color).

The main limitation of the above mentioned methods is that they only recover the most correlated (joint) linear subspace of the data, ignoring the individual components of the data. This is alleviated by recent methods such as

the Joint and Individual Variation Explained (JIVE) [20], the Common Orthogonal Basis Extraction (COBE) [41] and Robust Correlated and Individual Component Analysis (RCICA) [25]. Except for the recently proposed RCICA, the COBE and JIVE methods rely on least squares error minimisation and thus they are prone to gross errors and outliers [15], making the estimated components to be arbitrarily away from the true ones. Hence, their applicability to analyse visual data captured in unconstrained conditions (i.e., in-the-wild) is rather limited. On the other hand, the RCICA, even though is not sensitive to gross errors, it can be strictly applied in two sets of data only.

In this paper, we are concerned with the problem of recovering the common and individual components from an arbitrary number of (visual) data sets, captured in-the-wild and thus contaminated by gross non-Gaussian noise. To this end, we propose robust alternatives of the JIVE (Fig. 1). The proposed robust JIVE decomposes the data into three terms: a low-rank matrix that captures the joint variation across data sets, low-rank matrices accounting for structured variation individual to each data set, and a sparse matrix collecting the gross errors.

Arguably the human face is the most suitable object for demonstrating the effectiveness of the proposed method. This is due to the fact that face images convey rich information which can be perceived as a superposition of components associated with attributes, such as facial identity, expression, age etc. For instance, a set of images depicting expressive faces consists of components that are shared across all images (joint components) and imparts to the depicted objects the properties of human faces. Besides joint components, an expressive face consists of individual components that are related to different expressions. Such individual components can be expression-specific deformation of face, i.e., deformations around lips and eye in case of smiles. Similarly, a set of images depicting faces in different ages can be seen as a superposition of joint components that are invariant to the age and age-specific (individual) components that are individual to each age group (e.g., wrinkles).

The above observations motivate us to demonstrate the applicability of the proposed decomposition into two problems involving facial images captured under in-the-wild conditions, namely Facial Expression Synthesis and Face Age Progression. In addition, we selected the face analysis as the application area as (a) face is one of the very limited objects which has so many attributes (e.g., age, expression, and identity) and (b) there is a huge line of research on both facial expression synthesis and face age progression [16, 36, 8, 27, 35, 38], but there is no any method able to address both problems. The RJIVE is, to best of our knowledge, the first method that provides an unified solution on both problems. Summarising the contributions of the paper are as follows:

- A set of novel methods, referred to as Robust JIVE (RJIVE), for robust recovering of joint and individual components from a arbitrary number of data sets, is introduced. Furthermore, two efficient algorithms for the RJIVE have been developed.
- Based on the learnt joint and individual components, a suitable optimization problem that extract the corresponding modes of variation of unseen test samples, is proposed.
- We demonstrate the applicability of the proposed methods in two challenging computer vision tasks namely facial expression synthesis and face age progression in-the-wild.

2. Robust JIVE

Consider M data sets $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$, with $\mathbf{x}_j^{(i)} \in \mathbb{R}^{d^{(i)}}$, $j = 1, \ldots, J$ being a vectorized (visual) data sample[1]. In practice, visual data are often contaminated by gross, sparse non-Gaussian errors such as pixels corruption and partial image occlusions. The goal of the RJIVE is to robustly recover the joint components which are shared across all data sets as well as the components which are deemed individual for each one of those data sets. That is:

$$\mathbf{X} = \mathbf{J} + \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T + \mathbf{E}, \tag{1}$$

where $\mathbf{X} = \left[\mathbf{X}^{(1)^T}, \cdots, \mathbf{X}^{(M)^T}\right]^T \in \mathbb{R}^{q \times J}$, $\mathbf{J} = \left[\mathbf{J}^{(1)^T}, \cdots, \mathbf{J}^{(M)^T}\right]^T \in \mathbb{R}^{q \times J}$, $\{\mathbf{A}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$, $q = d^{(1)} + \cdots + d^{(M)}$, are low-rank matrices capturing the joint and individual variations, respectively and $\mathbf{E} \in \mathbb{R}^{q \times J}$ denotes the error matrix accounting for the gross, but sparse non-Gaussian noise. In order to ensure the identifiability of (1), the joint and common components should be mutual incoherent, i.e., $\{\mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^{M}$. A natural estimator accounting for the sparsity of the error matrix $\mathbf{E}$, is to minimize the number of the nonzero entries of $\mathbf{E}$ measured by the ℓ_0-quasi norm [4]. However, the minimization of the ℓ_0 norm is NP-hard due to its discrete nature [23]. This problem is typically addressed by adopting the ℓ_1 norm as the convex surrogate of the ℓ_0 norm [7]. Thus, the joint

[1] *Notation:* Throughout the paper, scalars are denoted by lower-case letters, vectors (matrices) are denoted by lower-case (upper-case) boldface letters i.e., $\mathbf{x}$, ($\mathbf{X}$). $\mathbf{I}$ denotes the identity matrix. The j-th column of $\mathbf{X}$ is denoted by $\mathbf{x}_j$. The ℓ_1 norm of $\mathbf{x}$ is defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$. The matrix ℓ_1 norm is defined as $\|\mathbf{X}\|_1 = \sum_i \sum_j |x_{ij}|$, where $|\cdot|$ denotes the absolute value operator. The Frobenius norm is defined as $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$, and the nuclear norm of $\mathbf{X}$ (i.e., the sum of singular values of a matrix) is denoted by $\|\mathbf{X}\|_*$.

and individual components as well as the sparse error are recovered by solving the following constrained non-linear optimization problem:

$$\min_{\mathbf{J},\{\mathbf{A}^{(i)}\}_{i=1}^{M}} \quad \left\|\mathbf{X}-\mathbf{J}-\left[\mathbf{A}^{(1)^T},\cdots,\mathbf{A}^{(M)^T}\right]^T\right\|_1. \\ \text{s.t.} \quad \operatorname{rank}(\mathbf{J})=r, \{\operatorname{rank}(\mathbf{A}^{(i)})=r^{(i)}, \mathbf{J}\mathbf{A}^{(i)^T}=\mathbf{0}\}_{i=1}^{M} \tag{2}$$

Clearly, (2) is a robust extension to JIVE [20] and requires an estimation for the rank of both joint and individual components. However, in practice those $(M+1)$ values are unknown and difficult to estimate. To alleviate this issue, we propose a variant of (2) which is able to determine the optimal ranks of individual components directly. By assuming that the actual ranks of individual components are upper bounded i.e., $\{\operatorname{rank}(\mathbf{A}^{(i)}) \leq K^{(i)}\}_{i=1}^{M}$, problem (2) is relaxed to the following one:

$$\begin{aligned}\min_{\mathbf{J},\{\mathbf{A}^{(i)}\}_{i=1}^{M}} \quad & \lambda\left\|\mathbf{X}-\mathbf{J}-\left[\mathbf{A}^{(1)^T},\cdots,\mathbf{A}^{(M)^T}\right]^T\right\|_1 \\ & +\sum_{i=1}^{M}\left\|\mathbf{A}^{(i)}\right\|_*, \\ \text{s.t.} \quad & \operatorname{rank}(\mathbf{J})=r, \{\mathbf{J}\mathbf{A}^{(i)^T}=\mathbf{0}\}_{i=1}^{M}\end{aligned} \tag{3}$$

where the nuclear norm acts as a convex surrogate of the rank function [9] and $\lambda>0$ is a regularizer.

2.1. Optimization Algorithms

In this section, algorithms for solving (4) and (18) are introduced.

2.1.1 Solving problem (3)

Problem (3), is difficult to be solved due to the presence of the orthogonality constraints $\{\mathbf{J}\mathbf{A}^{(i)^T}=\mathbf{0}\}_{i=1}^{M}$ and the non-differentiable but convex norms. The Alternating-Direction Method of Multipliers (ADMM) [2] has been proven to be very efficient for solving problems including the nuclear- and ℓ_1- norms [25, 30, 10, 40]. To this end, (3) is reformulated to the following separable one:

$$\begin{aligned}\min_{\mathbf{J},\{\mathbf{A}^{(i)},\mathbf{R}^{(i)}\}_{i=1}^{M},\mathbf{E}} \quad & \sum_{i=1}^{M}\left\|\mathbf{R}^{(i)}\right\|_* + \lambda\left\|\mathbf{E}\right\|_1, \\ \text{s.t.} \quad & \mathbf{X}=\mathbf{J}+\left[\mathbf{A}^{(1)^T},\cdots,\mathbf{A}^{(M)^T}\right]^T+\mathbf{E}, \\ & \operatorname{rank}(\mathbf{J})=r, \{\mathbf{R}^{(i)}=\mathbf{A}^{(i)}, \mathbf{J}\mathbf{A}^{(i)^T}=\mathbf{0}\}_{i=1}^{M}\end{aligned} \tag{4}$$

where $\{\mathbf{R}^{(i)}\in\mathbb{R}^{d^{(i)}\times J}\}_{i=1}^{M}$, $\{\mathbf{R}^{(i)}=\mathbf{A}^{(i)}\}_{i=1}^{M}$ are auxiliary variables and the corresponding constraints, respectively. To solve (4), the corresponding augmented Lagrangian function is defined as:

$$\begin{aligned}\mathcal{L}(\mathcal{B},\mathcal{V}) = & \sum_{i=1}^{M}\left\|\mathbf{R}^{(i)}\right\|_* + \lambda\left\|\mathbf{E}\right\|_1 - \frac{1}{2\mu}\left\|\mathbf{F}\right\|_F^2 \\ & +\frac{\mu}{2}\left\|\mathbf{X}-\mathbf{J}-\left[\mathbf{A}^{(1)^T},\cdots,\mathbf{A}^{(M)^T}\right]^T-\mathbf{E}+\frac{\mathbf{F}}{\mu}\right\|_F^2 \\ & +\sum_{i=1}^{M}\left(\frac{\mu}{2}\left\|\mathbf{R}^{(i)}-\mathbf{A}^{(i)}+\frac{\mathbf{Y}^{(i)}}{\mu}\right\|_F^2-\frac{1}{2\mu}\left\|\mathbf{Y}^{(i)}\right\|_F^2\right),\end{aligned} \tag{5}$$

where $\mathcal{B}=\{\mathbf{J},\{\mathbf{A}^{(i)},\mathbf{R}^{(i)}\}_{i=1}^{M},\mathbf{E}\}$ is the set of primal variables, $\mathcal{V}=\{\mathbf{F},\{\mathbf{Y}^{(i)}\}_{i=1}^{M}\}$ is the set of Lagrange multipliers related to the equality constraints in (4), and $\mu>0$ is a parameter. Then, by employing the ADMM, (5) is minimized with respect to each variable that belongs to $\mathcal{B}$ in an alternating fashion and finally the Lagrange multipliers $\mathcal{V}$ are updated. Let t be the iteration index. For notation convenience, $\mathcal{L}(\mathbf{J})$ will be used instead of $\mathcal{L}(\mathbf{J},\{\mathbf{A}^{(i)},\mathbf{R}^{(i)},\mathbf{Y}^{(i)}\}_{i=1}^{M},\mathbf{E},\mathbf{F})$ when all the variables except $\mathbf{J}$ are kept fixed. Therefore, given the primal variables $\mathcal{B}_t$, Lagrange multipliers $\mathcal{V}_t$ and parameter μ_t, the ADMM-based solver updates all the variables iteratively by solving the following optimization sub-problems:

J-subproblem:

$$\mathbf{J}_{t+1}=\min_{\mathbf{J}}\left\|\mathbf{X}-\mathbf{J}-\left[\mathbf{A}_t^{(1)^T},\cdots,\mathbf{A}_t^{(M)^T}\right]^T-\mathbf{E}_t+\frac{\mathbf{F}_t}{\mu_t}\right\|_F^2, \\ \text{s.t.}\,\operatorname{rank}(\mathbf{J})=r. \tag{6}$$

The rank constrained least-squares problem (6) has the following closed-form solution:

$$\mathbf{J}_{t+1}=\mathcal{Q}_{\mathbf{r}}\left[\mathbf{X}-\left[\mathbf{A}_t^{(1)^T},\cdots,\mathbf{A}_t^{(M)^T}\right]^T-\mathbf{E}_t+\frac{\mathbf{F}_t}{\mu_t}\right]. \tag{7}$$

The $\mathcal{Q}_{\mathbf{r}}$ operator is defined for any matrix $\mathbf{D}$ with $\mathbf{D}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ as $\mathcal{Q}_r[\mathbf{D}]=\left[\mathbf{U}(:,1:r)\mathbf{\Sigma}(1:r,1:r)\mathbf{V}(:,1:r)^T\right]$.

$\mathbf{A}^{(i)}$-subproblem:

$$\mathbf{A}_{t+1}^{(i)}=\min_{\mathbf{A}^{(i)}}\left\|\mathbf{X}^{(i)}-\mathbf{J}_{t+1}^{(i)}-\mathbf{A}^{(i)}-\mathbf{E}_t^{(i)}+\frac{\mathbf{F}_t^{(i)}}{\mu_t}\right\|_F^2 \\ +\left\|\mathbf{R}^{(i)}-\mathbf{A}^{(i)}+\frac{\mathbf{Y}_t^{(i)}}{\mu_t}\right\|_F^2, \text{s.t.}\quad \mathbf{J}_{t+1}\mathbf{A}^{(i)^T}=\mathbf{0}. \tag{8}$$

Based on the Proposition 1, the solution of (8) is given by:

$$\mathbf{A}_{t+1}^{(i)}=\frac{(\mathbf{X}^{(i)}-\mathbf{J}_{t+1}^{(i)}-\mathbf{E}_t^{(i)}+\mathbf{R}_t^{(i)}+\frac{\mathbf{F}_t^{(i)}+\mathbf{Y}_t^{(i)}}{\mu_t})\mathbf{P}}{2}, \tag{9}$$

Algorithm 1: ADMM solver of (4) (NN-ℓ_1-RJIVE).

Input : Data $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$. Rank of joint component r. Parameter ρ.
Output : Joint component $\mathbf{J}$, individual components $\{\mathbf{A}^{(i)}\}_{i=1}^{M}$
Initialize: Set $\mathbf{J}_0$, $\{\mathbf{A}_0^{(i)}, \mathbf{R}_0^{(i)}, \mathbf{Y}_0^{(i)}\}_{i=1}^{M}$, $\mathbf{E}_0$, $\mathbf{F}_0$ to zero matrices, $t = 0$, $\mu_0 > 0$.

$\mathbf{X} = \left[\mathbf{X}^{(1)^T}, \cdots, \mathbf{X}^{(M)^T}\right]^T$;
while *not converged* **do**
  Compute $\mathbf{J}_{t+1}$ by (7);
  for $i = 1 : M$ **do**
    Compute $\mathbf{A}_{t+1}^{(i)}$, $\mathbf{R}_{t+1}^{(i)}$, $\mathbf{Y}_{t+1}^{(i)}$ by (9), (12), (16);
  end
  Compute $\mathbf{E}_{t+1}$ by (14);
  Compute $\mathbf{F}_{t+1}$ and μ_t by (15), (17);
  $t = t + 1$;
end

where $\mathbf{P}$ is a projection to the orthogonal complement of $\mathbf{V}(:, 1:r)\mathbf{V}(:, 1:r)^T$, namely $\mathbf{P} = \mathbf{I} - \mathbf{V}(:, 1:r)\mathbf{V}(:, 1:r)^T$ and $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{X} - [\mathbf{A}_t^{(1)^T}, \cdots, \mathbf{A}_t^{(M)^T}]^T - \mathbf{E}_t + \mathbf{F}_t/\mu_t$ is the Singular Value Decomposition (SVD).

Proposition 1 *The constrained optimization problem:*

$$\min_{\mathbf{Z}} \|\mathbf{B} - \mathbf{Z}\|_F^2, \quad s.t. \quad \mathbf{B}\mathbf{Z}^T = \mathbf{0} \tag{10}$$

has a closed-form solution given by $\mathbf{Z} = \mathbf{B}\left(\mathbf{I} - \mathbf{V}\mathbf{V}^T\right)$, $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

$\mathbf{R}^{(i)}$-subproblem:

$$\mathbf{R}_{t+1}^{(i)} = \min_{\mathbf{R}^{(i)}} \left\|\mathbf{R}^{(i)}\right\|_* + \frac{\mu_t}{2}\left\|\mathbf{R}^{(i)} - \mathbf{A}_{t+1}^{(i)} + \frac{\mathbf{Y}_t^{(i)}}{\mu_t}\right\|_F^2. \tag{11}$$

The solution of the nuclear-norm regularized least-squares subproblem (11), is obtained by applying the Singular Value Thresholding (SVT) [3] operator $\mathcal{D}_\tau$:

$$\mathbf{R}_{t+1}^{(i)} = \mathcal{D}_{1/\mu_t}\left[\mathbf{A}_{t+1}^{(i)} - \frac{\mathbf{Y}_t^{(i)}}{\mu_t}\right]. \tag{12}$$

For any matrix $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ the SVT operator is defined as $\mathcal{D}_\tau = \mathbf{U}\mathcal{S}_\tau\mathbf{V}^T$. The shrinkage operator $\mathcal{S}_\tau$ [4] is defined element-wise as $\mathcal{S}_\tau[\sigma] = \text{sgn}(\sigma)\max(|\sigma| - \tau, 0)$.

E-subproblem:

$$\mathbf{E}_{t+1} = \min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 + \frac{\mu_t}{2}\left\|\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T - \mathbf{E} + \frac{\mathbf{F}_t}{\mu_t}\right\|_F^2 \tag{13}$$

The closed-form solution of (13) is given by the shrinkage operator:

$$\mathbf{E}_{t+1} = \mathcal{S}_{\frac{\lambda}{\mu_t}}\left[\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T + \frac{\mathbf{F}_t}{\mu_t}\right]. \tag{14}$$

Update Lagrange multipliers and parameter μ:

$$\mathbf{F}_{t+1} = \mathbf{F}_t + \mu_t\Big(\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T - \mathbf{E}_{t+1}\Big). \tag{15}$$

$$\mathbf{Y}_{t+1}^{(i)} = \mathbf{Y}_t^{(i)} + \mu_t(\mathbf{R}_{t+1}^{(i)} - \mathbf{A}_{t+1}^{(i)}), \quad i = 1, \ldots, M \tag{16}$$

$$\mu_{t+1} = \min(\rho\mu_t, 10^7) \tag{17}$$

The ADMM solver of (4) is outlined in Alg. 1. The dominant cost of Algorithm is mainly associated to the SVD. Therefore, the complexity of each iteration is $\mathcal{O}(\max(q^2J, qJ^2))$. Alg. 1 is terminated when $\left\|\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T - \mathbf{E}_{t+1}\right\|_F^2 / \|\mathbf{X}\|_F^2$ is less than a predefined threshold ϵ_1 or the number of iterations reach a maximum value.

2.1.2 Solving problem (2)

To solve problem (2) via ADMM, we firstly reformulate it as:

$$\begin{aligned} &\min_{\mathbf{J}, \{\mathbf{A}^{(i)}\}_{i=1}^{M}, \mathbf{E}} \quad \|\mathbf{E}\|_1, \\ &\text{s.t.} \quad \mathbf{X} = \mathbf{J} + \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T + \mathbf{E}, \\ &\qquad \text{rank}(\mathbf{J}) = r, \{\text{rank}(\mathbf{A}^{(i)}) = r^{(i)}, \mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^{M} \end{aligned} \tag{18}$$

The corresponding augmented Lagrangian function is given by:

$$\mathcal{L}(\mathbf{J}, \{\mathbf{A}^{(i)}\}_{i=1}^{M}, \mathbf{E}, \mathbf{L}) = \|\mathbf{E}\|_1 - \frac{1}{2\mu}\|\mathbf{L}\|_F^2 + \frac{\mu}{2}\left\|\mathbf{X} - \mathbf{J} - \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T - \mathbf{E} + \frac{\mathbf{L}}{\mu}\right\|_F^2. \tag{19}$$

The ADMM solver of (18) is wrapped up in Alg. 2 while its derivation is similar to that of Alg. 1. Furthermore, Alg. 2 has the same complexity and convergence criterion as Alg. 1.

3. RJIVE-Based Reconstruction

Having recovering the individual and common components of the M data sets, we can effectively use them in order to extract the joint and individual mode of variations of a test sample. For instance, the components recovered by applying the RJIVE in a set of facial images

Algorithm 2: ADMM solver for (18) (ℓ_1-RJIVE).

Input : Data $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$. Rank of joint component r. Ranks of individual components $\{r^{(i)}\}_{i=1}^{M}$. Parameter ρ.

Output : Joint component $\mathbf{J}$, individual components $\{\mathbf{A}^{(i)}\}_{i=1}^{M}$

Initialize: Set $\mathbf{J}_0$, $\{\mathbf{A}_0^{(i)}\}_{i=1}^{M}$, $\mathbf{E}_0$, $\mathbf{L}_0$ to zero matrices, $t = 0$, $\mu_0 > 0$.

$\mathbf{X} = \left[\mathbf{X}^{(1)^T}, \cdots, \mathbf{X}^{(M)^T}\right]^T$;

while *not converged* **do**

  $\mathbf{M} = \mathbf{X} - \left[\mathbf{A}_t^{(1)^T}, \cdots, \mathbf{A}_t^{(M)^T}\right]^T - \mathbf{E}_t + \mu_t^{-1}\mathbf{L}_t$;

  $\mathbf{J}_{t+1} = \mathcal{Q}_r[\mathbf{M}]$, $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{svd}(\mathbf{M})$;

  $\mathbf{P} = \mathbf{I} - \mathbf{V}(:, 1:r)\mathbf{V}(:, 1:r)^T$;

  for $i = 1 : M$ **do**

    $\mathbf{A}_{t+1}^{(i)} = \mathcal{Q}_{r^{(i)}}\left[\left(\mathbf{X}^{(i)} - \mathbf{J}_{t+1}^{(i)} - \mathbf{E}_t^{(i)} + \mu_t^{-1}\mathbf{L}_t^{(i)}\right)\mathbf{P}\right]$

  end

  $\mathbf{E} = \mathcal{S}_{\frac{1}{\mu_t}}\left[\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T - \mu_t^{-1}\mathbf{L}\right]$

  $\mathbf{L}_{t+1} = \mathbf{L}_t + \mu_t\left(\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T - \mathbf{E}_{t+1}\right)$;

  $\mu_{t+1} = \min(\rho \cdot \mu_t, 10^7)$;

  $t = t + 1$;

end

of different expressions can be utilized in order synthesize the 'Happy' and 'Surprise' expression of a 'Neutral' subject. Let $\{\mathbf{U}_{\mathbf{J}^{(i)}} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{J}}^{(i)}}, \mathbf{U}_{\mathbf{A}^{(i)}} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{A}}^{(i)}}, \}_{i=1}^{M}$ be orthonormal bases extracted by applying the SVD onto $\{\mathbf{J}^{(i)}, \mathbf{A}^{(i)}\}_{i=1}^{M}$ and $\mathbf{t} \in \mathbb{R}^{d^{(i)} \times 1}$ a new sample. The corresponding i-th sets of variables or modality of $\mathbf{t}$ is reconstructed by solving the following constrained optimization problem:

$$\begin{aligned} \min_{\{\mathbf{c}^{(n)}, \mathbf{v}^{(n)}\}_{n=1}^{2}, \mathbf{y} \geq \mathbf{0}} \quad & \sum_{n=1}^{2} \left\|\mathbf{v}^{(n)}\right\|_1 + \theta \left\|\mathbf{e}\right\|_1, \\ \text{s.t.} \quad & \{\mathbf{v}^{(n)} = \mathbf{c}^{(n)}\}_{n=1}^{2} \\ & \mathbf{t} = \mathbf{U}_{\mathbf{J}^{(i)}}\mathbf{c}^{(1)} + \mathbf{U}_{\mathbf{A}^{(i)}}\mathbf{c}^{(2)} + \mathbf{e} \\ & \mathbf{y} = \mathbf{U}_{\mathbf{J}^{(i)}}\mathbf{c}^{(1)} + \mathbf{U}_{\mathbf{A}^{(i)}}\mathbf{c}^{(2)} \end{aligned} \tag{20}$$

where θ is a positive parameter that balances the norms, $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$ are auxiliary variables, $\mathbf{y}$ corresponds to the non-negative clean reconstruction, and $\mathbf{e}$ is an error term.

4. Experimental Evaluation

4.1. Synthetic

In this section, the ability of RJIVE to robustly recover the joint and individual components of synthetic data corrupted by sparse non-Gaussian noise, is tested. To this end, sets of matrices $\{\mathbf{X}^{(i)} = \mathbf{J}_*^{(i)} + \mathbf{A}_*^{(i)} + \mathbf{E}_*^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{2}$ of varying dimensions were generated. In more detail, a rank-r joint component $\mathbf{J}_* \in \mathbb{R}^{(q=d^{(1)}+d^{(2)}) \times J}$ was created from a random matrix $\mathbf{X} = [\mathbf{X}^{(1)^T}, \mathbf{X}^{(2)^T}]^T \in \mathbb{R}^{q \times J}$. Next, the orthogonal to $\mathbf{J}$ rank-$r^{(1)}$, $r^{(2)}$ common components $\mathbf{A}_*^{(1)}$ and $\mathbf{A}_*^{(2)}$ were computed by $[\mathbf{A}_*^{(1)^T}, \mathbf{A}_*^{(2)^T}]^T = (\mathbf{X} - \mathbf{J}_*)(\mathbf{I} - \mathbf{V}\mathbf{V}^T)$, where $\mathbf{V}$ was formed from the first r columns of the row space of $\mathbf{X}$. $\mathbf{E}_*^{(i)}$ is a sparse error matrix with 20% non-zero entries being sampled independently from $\mathcal{N}(0, 1)$.

The Relative Reconstruction Error (RRE) of the joint and individual components achieved by both ℓ_1-RJIVE and Nuclear-Norm regularized (NN-ℓ_1-RJIVE) for a varying number of dimensions, joint and individual ranks, are reported in Table 1. The corresponding RRE obtained by JIVE [20], and COBE [41] are also presented. As it can be seen, the proposed methods accurately recovered both the joint and individual components. It is worth mentioning that the NN-ℓ_1-RJIVE successfully recovered all components by utilizing only the true rank of the joint component. In contrast, all the other methods require knowledge regarding the true rank for both joint and individual components. Based on the performance of NN-ℓ_1-RJIVE on the synthetic data, we decided to exploit it in the experiments described bellow and referred to as RJIVE hereafter.

4.2. Facial Expression Synthesis (FES) In-The-Wild

In this section, we investigate the ability of the RJIVE to synthesize a set of different expressions of a given facial image. Consider M data sets where each one contains images of different subjects that depict a specific expression. In order to effectively recover the joint and common components, the faces of each data set should be put in correspondence. Thus, their $N = 68$ facial landmark points are localized using the detector from [1] trained with images provided from 300-W challenge [34, 29, 33] and subsequently employed to compute a mean reference shape. Then, the faces of each data set are warped into corresponding reference shape by using the piecewise affine warp function $\mathcal{W}(\cdot)$ [21, 6]. After applying the RJIVE on the warped data sets, the recovered components can be used for synthesizing M different expressions of an unseen subject. To do that, the new (unseen) facial image is warped to reference frame corresponds to expression that we want to synthesize and subsequently is given as input to the solver of (20).

The performance of RJIVE in FES task is assessed

Table 1. Quantitative recovering results produced by JIVE [20], COBE [41], ℓ_1-RJIVE (18), and NN-ℓ_1-RJIVE (4).

$\left(d^{(1)}, d^{(2)}, J, r, r^{(1)}, r^{(2)}\right)$	Method	$\left\{\frac{\left\|\mathbf{J}_*^{(i)}-\mathbf{J}^{(i)}\right\|_F^2}{\left\|\mathbf{J}_*^{(i)}\right\|_F^2}\right\}_{i=1}^2$	$\left\{\frac{\left\|\mathbf{A}_*^{(i)}-\mathbf{A}^{(i)}\right\|_F^2}{\left\|\mathbf{A}_*^{(i)}\right\|_F^2}\right\}_{i=1}^2$	Time (in CPU seconds)
$(500, 500, 500, 5, 10, 10)$	COBE	3.6403	1.0956	0.14
	JIVE	0.5424	0.93488	26
	ℓ_1-RJIVE	$5.5628e-08$	$3.5073e-08$	13
	NN-ℓ_1-RJIVE	$1.3338e-08$	$1.1897e-08$	13
$(1000, 1000, 1000, 10, 20, 20)$	COBE	5.2902	1.0945	3.3
	JIVE	0.83977	1.481	58
	ℓ_1-RJIVE	$8.9035e-08$	$5.8665e-08$	27
	NN-ℓ_1-RJIVE	$9.78e-08$	$8.9923e-08$	46
$(2000, 2000, 2000, 20, 40, 40)$	COBE	7.3973	1.1808	30
	JIVE	1.3961	2.1977	390
	ℓ_1-RJIVE	$2.427e-07$	$2.3743e-06$	310
	NN-ℓ_1-RJIVE	$2.2969e-07$	$2.0574e-07$	351

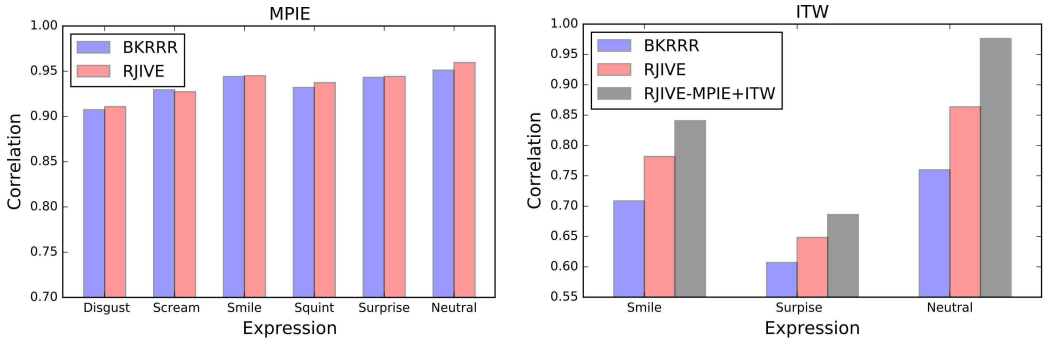
Figure 2. Synthesized expressions of MPIE's subject '014' produced by the BKRRR and RJIVE methods.

by conducting inner- and cross-databases experiments on MPIE [11] and in-the-wild facial images collected from the internet (ITW). The synthesized expressions obtained by RJIVE[2] are compared to those obtained by the state-of-the-art BKRRR [13] method. In particular, the BKRRR is a regression-based method that learns a mapping from the 'Neutral' expression to the target ones. Then, given the 'Neutral' face of an unseen subject, new expressions are synthesized by employing the corresponding learnt regression functions. The performance of the compared methods is measured by computing the correlation between the vectorized forms of true images and the reconstructed ones. In the first experiment, 534 frontal images of MPIE database that depict 89 subjects under six expressions (i.e., 'Neutral', 'Scream', 'Squint', 'Surprise', 'Smile', 'Disgust') were employed to train both RJIVE and BKRRR. Then, all expressions of 58 unseen subjects from the same database were synthesized by using their images correspond to 'Neutral' expressions. In Fig. 3(a) the average correlations obtained by the compared methods for the different expressions are visualized. As it can be seen the proposed RJIVE method achieves the same accuracy to BKRRR without learning any kind of mappings between the different expressions of the same subject. Specifically, the RJIVE extracts only the individual components of each expression and the common one.

Furthermore, we collected from the internet 180 images depicting 60 subjects with 'Surprise', 'Smile', and 'Neutral' expressions (three images for each subject). Then, all the expressions were generated by employing the 'Neutral' images and the BKRRR and RJIVE[3] methods trained on MPIE. Figure 3(b) depicts the obtained correlations for each subject. Clearly, the RJIVE outperforms the BKRRR. Compared to the previous experiment, there is a drop in performance for both methods. This is attributed to the fact that the methods were trained by employing only images captured under controlled condition. Thus, synthesizing expressions of in-the-wild images is a very difficult task. In order to alleviate this problem we can augment the training set with in-the-wild images. Although the RJIVE can be trained from in-the-wild images of different subjects, this is not the case of BKRRR, which requires the correspondence of expressions across the training subjects. Collecting in-the-wild images of same subjects under different expressions is a very tedious task. In order to improve the performance of RJIVE, we augmented the training set with another 1200 images from WWB database [22] (400 images for each expression). As it can be observed in Fig. 3(b), the in-the-wild train set improved the accuracy of RJIVE in the ITW dataset. Figure 4 depicts examples synthesized in-the-wild expressions produced by the RJIVE. The images from the 'Input' column were given as input to the RJIVE and

[2] $(r, \lambda, W_{\mathbf{J}}^{(i)}, W_{\mathbf{A}}^{(i)}, \theta, \epsilon_1) = (20, \frac{1}{\sqrt{\max(q,J)}}, 70, 70, 0.03, 10^{-5})$.

[3] $(r, \lambda, W_{\mathbf{J}}^{(i)}, W_{\mathbf{A}}^{(i)}, \theta, \epsilon_1) = (150, \frac{1}{\sqrt{\max(q,J)}}, 300, 300, 0.03, 10^{-5})$.



(a) (b)

Figure 3. Mean average correlation achieved by JIVE and BKRRR methods on (a) MPIE and (b) ITW databases.

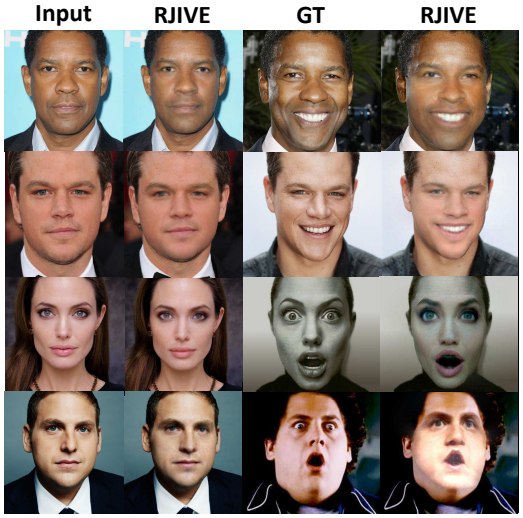


Figure 4. Synthesized in-the-wild expressions produced by the RJIVE method.

subsequently the synthesized expressions were warped and fused with the actual images [26]. Clearly, the produced expressions are characterized by high quality of both expression and identity information. It is worth mentioning that RJIVE synthesize almost perfectly the input images without using any kind of information about the depicted subject.

4.3. Face Age Progression In-The-Wild

Face age progression consists in synthesizing plausible faces of subjects at different ages. It is considered as a very challenging task due to the fact that the face is a highly deformable object and its appearance drastically changes under different illumination conditions, expressions, and poses. Various databases that contain faces at different ages have been collected in the last couple of years [5, 28]. Although these databases contain huge number of images, they have some limitations including limited images for each subject that cover a narrow range of ages and noisy age labels, since most of them have been collected by employing automatic procedures (crawlers). Recently, the AgeDB database that overcomes the aforementioned problems have been proposed in [39]. In order to train the RJIVE, the AgeDB was divided into $M = 10$ age groups (700 images per group): $0-3$, $4-7$, $8-15$, $16-20$, $21-30$, $31-40$, $41-50$, $51-60$, $61-70$, and $71-100$. Then, following the same procedure as in FES task, the RJIVE was employed to extract the joint and common components from the warped images. The performance of RJIVE in face age progression in-the-wild is qualitatively assessed conducting experiments on images from the FG-NET database [19]. To this end, we compare the performance of RJIVE with the Illumination Aware Age Progression (IAAP) method [16], Coupled Dictionary Learning (CDL) method [36], Deep Aging with Restricted Boltzmann Machines (DARB) method [8], CG [27], and Recurrent Face Aging (RFA) method [38]. In Fig. 5 progressed images produced by the compared methods are depicted. Note, that all the progressed faces have been warped back and fused with the actual ones.

The performance of the RJIVE[4] is also quantitatively assessed by conducting age-invariant face verification experiments. Following the successfully used verification protocol of the LFW database [14], we propose four new age-invariant face verification protocols based on the proposed AgeDB database. Each one of the protocols was created by splitting the AgeDB database into 10 folds, with each fold consisting of 300 intra-class pairs and 300 inter-class pairs. The essential difference between these protocols is that in each protocol the age difference of each pair's faces is equal to a predefined value i.e., $\{5$ ages, 10 ages, 20 ages, 30 ages$\}$.

In order to assess the performance of RJIVE, the following procedure was performed. For each fold of a specific protocol the training images were split into $M = 10$ age-groups and subsequently the RJIVE was employed on their warped version in order to extract the joint and individual components. All images of each training pair were

[4] $(r, \lambda, W_{\mathbf{J}}^{(i)}, W_{\mathbf{A}}^{(i)}, \theta, \epsilon_1) = (300, \frac{1}{\sqrt{\max(q,J)}}, 600, 600, 0.03, 10^{-5})$.

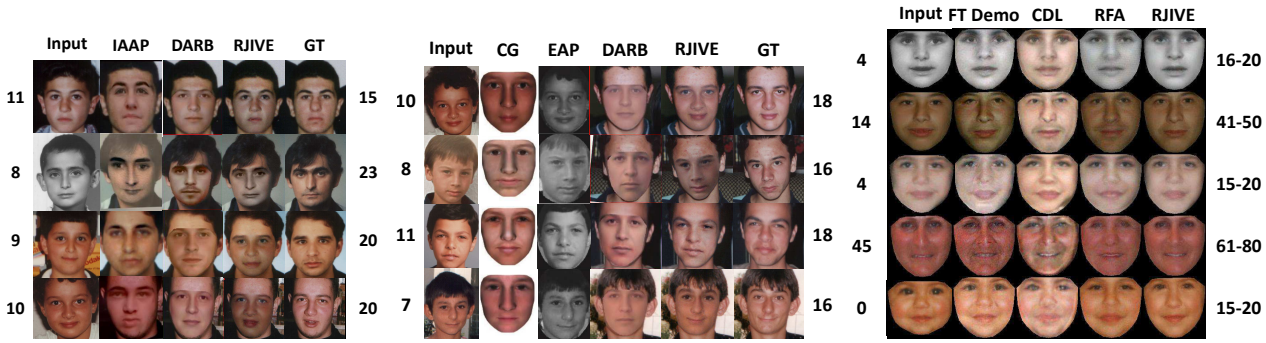


Figure 5. Progressed faces produced by the compared methods on the FG-NET database.

then progressed into $M = 10$ age groups resulting into 10 new pairs. As we wanted to represent each pair by using a single feature, gradients orientations were extracted from the corresponding images and subsequently the mean value of their cosine difference was employed as the pair's feature [31, 32]. M Support Vector Machines (SVMs) were trained by utilizing the extracted features and their scores were lately fused by using an SVM.

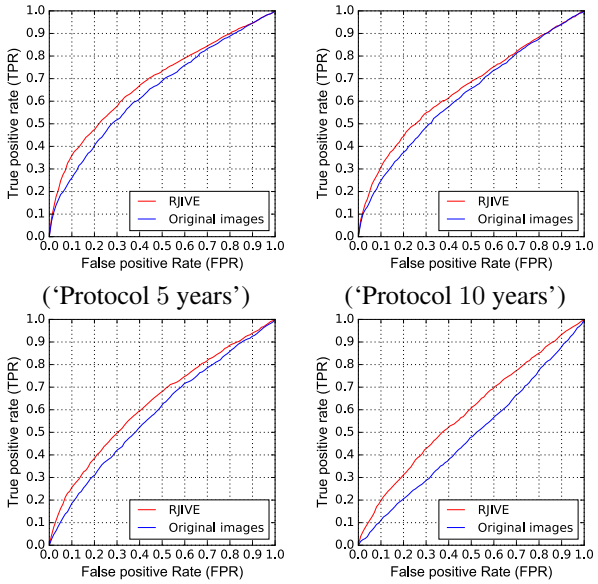
Table 2. Mean AUC and Accuracy on the proposed four protocols.

	RJIVE		Original Images	
Protocol	AUC	Accuracy	AUC	Accuracy
5 years	0.686	0.637	0.646	0.609
10 years	0.654	0.621	0.624	0.591
20 years	0.633	0.598	0.585	0.552
30 years	0.584	0.552	0.484	0.495

In Fig. 6, Receiver Operating Characteristic (ROC) curves computed based on the 10 folds of each one of the proposed protocols are depicted. The corresponding mean classification accuracy and Are Under Curve (AUC) are reported in Table 2. In order to assess the effect of progression, the results obtained by utilizing only the original images are also provided. Some interesting observations are drawn from the results. Firstly, the improvement in accuracy validates that the identity information of the face remains after the RJIVE-based progression. Furthermore, the improvement in accuracy is higher when the age difference of images of each pair is big enough. For instance, the improvement in accuracy in 'Protocol 30 years' is higher than the corresponding in 'Protocol 5 years'. Finally, the produced results justify that the problem of age-invariant face verification becomes more difficult when the age difference is very large (e.g., 30 years).

5. Conclusions

In this paper, the RJIVE along with its algorithmic framework for robust recovery of joint and individual variance among several visual data sets has been proposed. The performance of the RJIVE has been assessed on facial expression synthesis, and face age progression by conducting experiments on data sets captured under both constrained and in-the-wild conditions. The experimental results validate the effectiveness of the proposed RJIVE method over the state-of-the-art.



('Protocol 5 years') ('Protocol 10 years')

('Protocol 20 years') ('Protocol 30 years')

Figure 6. ROC curves of RJIVE on the proposed four protocols. 'Original images' corresponds to the results obtained by employing the actual images.

6. Acknowledgements

The work of S. Zafeiriou was partially funded by EPSRC project FACER2VM (EP/N007743/1). The work of C. Sagonas was partially funded by the FiDiPro program of Tekes (project number: 1849/31/2015), as well as by the European Community Horizon 2020 [H2020/2014-2020] under grant agreement no. 688520 (TeSLA). The work of Y. Panagakis was partially funded by the European Community Horizon 2020 under grant agreement no. 645094 (SEWA).

References

[1] J. Alabort-i-Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *Proceedings of the ACM International Conference on Multimedia, Open Source Software Competition*, pages 679–682, Orlando, FL, USA, November 2014.

[2] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[3] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[5] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 768–783. Springer, 2014.

[6] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 23(6):681–685, 2001.

[7] D. L. Donoho. For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

[8] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, 2016.

[9] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[10] C. Georgakis, Y. Panagakis, and M. Pantic. Dynamic behavior analysis via structured rank minimization. *International Journal of Computer Vision (IJCV)*, pages 1–25, 2017.

[11] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing (IMAVIS)*, 28(5):807–813, 2010.

[12] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[13] D. Huang and F. De la Torre. Bilinear kernel reduced rank regression for facial expression synthesis. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 364–377. Springer, 2010.

[14] G. B. Huang and E. Learned-Miller. Labeled faces in the wild: Updates and new reporting procedures. *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, pages 14–003, 2014.

[15] P. J. Huber. *Robust statistics*. Springer, 2011.

[16] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz. Illumination-aware age progression. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, 2014.

[17] A. Klami and S. Kaski. Probabilistic approach to detecting dependencies between data sets. *Neurocomputing*, 72(1):39–46, 2008.

[18] A. Klami, S. Virtanen, E. Leppäaho, and S. Kaski. Group factor analysis. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 26(9):2136–2147, 2015.

[19] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(4):442–455, 2002.

[20] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel. Joint and individual variation explained (jive) for integrated analysis of multiple data types. *The annals of applied statistics*, 7(1):523, 2013.

[21] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision (IJCV)*, 60(2):135–164, 2004.

[22] A. Mollahosseini, B. Hassani, M. J. Salvador, H. Abdollahi, D. Chan, and M. H. Mahoor. Facial expression recognition from world wild web. *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition, Workshops (CVPR-W)*, 2016.

[23] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[24] M. A. Nicolaou, V. Pavlovic, and M. Pantic. Dynamic probabilistic cca for analysis of affective behavior and fusion of continuous annotations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(7):1299–1311, 2014.

[25] Y. Panagakis, M. Nicolaou, S. Zafeiriou, and M. Pantic. Robust correlated and individual component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 38(8):1665–1678, 2015.

[26] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 313–318. ACM, 2003.

[27] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, volume 1, pages 387–394, 2006.

[28] R. Rothe, R. Timofte, and L. Van Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition, Workshops (CVPR-W)*, pages 10–15, 2015.

[29] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: database and results. *Image and Vision Computing (IMAVIS)*, 47:3–18, 2016.

[30] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic. RAPS: Robust and efficient automatic construction of person-specific deformable models. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, pages 1789–1796, 2014.

[31] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic. Robust statistical face frontalization. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 3871–3879, 2015.

[32] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic. Robust statistical frontalization of human and animal faces. *International Journal of Computer Vision (IJCV)*, 122(2):270–291, 2017.

[33] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of IEEE International Conference on Computer Vision (ICCV-W), Workshop on 300 Faces in-the-Wild Challenge (300-W)*, pages 397–403, Sydney, Australia, 2013.

[34] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. A semi-automatic methodology for facial landmark annotation. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pages 896–903, 2013.

[35] C.-T. Shen, W.-H. Lu, S.-W. Shih, and H.-Y. M. Liao. Exemplar-based age progression prediction in children faces. In *IEEE International Symposium on Multimedia (ISM)*, pages 123–128, 2011.

[36] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan. Personalized age progression with aging dictionary. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, pages 3970–3978, 2015.

[37] L. R. Tucker. An inter-battery method of factor analysis. *Psychometrika*, 23(2):111–136, 1958.

[38] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe. Recurrent face aging. *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, 2016.

[39] S. Zafeiriou, A. Papaioannou, S. Moschoglou, C. Sagonas, J. Deng, and I. Kotsia. AgeDB: A new database for age related face analysis: Verification, Progression and Beyond. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition, Workshops (CVPR-W)*, 2017.

[40] T. Zhang, K. Jia, C. Xu, Y. Ma, and N. Ahuja. Partial occlusion handling for visual tracking via robust part matching. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, pages 1258–1265, 2014.

[41] G. Zhou, A. Cichocki, Y. Zhang, and D. P. Mandic. Group component analysis for multiblock data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 11:2426–2439, 2015.