# Recovering Joint and Individual Components in Facial Data

Christos Sagonas, Evangelos Ververas, Yannis Panagakis, and Stefanos Zafeiriou, *Member, IEEE*

**Abstract**—A set of images depicting faces with different expressions or in various ages consists of components that are shared across all images (i.e., *joint* components) imparting to the depicted object the properties of human faces as well as *individual* components that are related to different expressions or age groups. Discovering the common (joint) and individual components in facial images is crucial for applications such as facial expression transfer and age progression. The problem is rather challenging when dealing with images captured in unconstrained conditions in the presence of sparse non-Gaussian errors of large magnitude (i.e., sparse gross errors or outliers) and contain missing data. In this paper, we investigate the use of a method recently introduced in statistics, the so-called Joint and Individual Variance Explained (JIVE) method, for the robust recovery of joint and individual components in visual facial data consisting of an arbitrary number of views. Since the JIVE is not robust to sparse gross errors, we propose alternatives, which are (1) robust to sparse gross, non-Gaussian noise, (2) able to automatically find the individual components rank, and (3) can handle missing data. We demonstrate the effectiveness of the proposed methods to several computer vision applications, namely facial expression synthesis and 2D and 3D face age progression 'in-the-wild'.

**Index Terms**—Low-Rank, Sparsity, Facial Expression Synthesis, Face Age Progression, Joint and Individual Components.

## 1 Introduction

Facial 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 (i.e., *joint* components) imparting to the depicted object 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 a face, i.e., deformations around lips and eyes 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 components that are individual to each age group (e.g., wrinkles). Consequently, being able to extract such joint and individual components from facial images is crucial for applications such as facial expression synthesis and age progression [1], [2], [3], [4], [5], [6], among other visual data analysis tasks.

Extracting the joint components among data has created a wealth of research in statistics, signal processing, and computer vision. Two mathematically similar but conceptually different models underlie the bulk of the methodologies. In particular, the Canonical Correlation Analysis (CCA) [7] and its variants e.g., [8], [9], have been proposed for extracting linear correlated components among two or more sets of variables. Similarly, inter-battery factor analysis [10] and its extensions e.g., [11], determines the common factors among two sets of variables. The main limitation of the aforementioned methods is that they only recover the most correlated linear subspace of the data, ignoring the individual components among the different views or datasets.

The above mentioned limitation is alleviated by recent methods such as the Joint and Individual Variation Explained (JIVE) [12], the Common Orthogonal Basis Extraction (COBE) [13], and the Robust Correlated and Individual Component Analysis (RCICA) [14], which are briefly described in Section 2.

Besides the rich structure in facial visual data, images are subject to various types of errors, distortions, and noise. Common dense distortions such as ambient noise or quantization noise are of small magnitude and it is natural to assume that they follow a Gaussian distribution of small variance. Methods such as the CCA and its variants, JIVE, and COBE are stable in the presence of Gaussian noise.

Apart from these small but dense noises, there are gross errors that are sparsely supported but of large or even unbounded magnitude, such as the salt-and-pepper noise in imaging devices, occlusions in facial images, registration errors, or errors due incorrect localization and tracking. These errors rarely follow a Gaussian distribution and due to their sparse nature (i.e., the number of errors is bounded below some constant) are collectively referred to as sparse gross errors or noise. Except for the most recent RCICA, the COBE and JIVE rely on least squares error minimization and thus they are prone to gross errors and outliers [15]. That is, the estimated components can be arbitrarily away from the true ones. Therefore, the problem of joint and individual components recovery is rather challenging when dealing with facial images and in general visual data captured under unconstrained (i.e., 'in-the-wild') conditions.

In this paper, we investigate the problem of recovering the joint and individual components from facial (and in general visual) data consisting of an arbitrary number of views, captured in-the-wild. Such data are therefore contaminated by sparse, gross, non-Gaussian noise and possibly contain missing values. To this end,

- *C. Sagonas is with Onfido, London WC2E 9LG, UK (email: ch.sagonas@gmail.com). This work was completed while C. Sagonas was at Imperial College London, UK.*
- *V. Ververas, Y. Panagakis, and S. Zafeiriou are with the Department of Computing, Imperial College London, SW7 2RH, London, UK.*
- *Y. Panagakis is also with the Department of Computer Science, Middlesex University London, UK.*

we propose robust alternatives to the JIVE (coined collectively as Robust-JIVE, RJIVE), where the components are estimated by employing the $\ell_1$-norm. The $\ell_1$-norm is suitable for robust estimation in the presence of sparse gross errors [15]. The contributions of the paper are summarized as follows:

- We propose a novel, general framework, the RJIVE in Section 3, for the robust recovering of joint and individual components from multi-view data in the presence of sparse gross errors and possibly missing values. The proposed RJIVE decomposes the data into three terms: a low-rank matrix that captures the joint variation across views, low-rank matrices accounting for structured variation individual to each view, and a sparse matrix collecting the sparse gross errors[1].
- In particular, the RJIVE consists of 4 different models, namely $\ell_1$-RJIVE, NN-$\ell_1$-RJIVE, SRJIVE, and RJIVE-M. In the $\ell_1$-RJIVE, the rank of both joint and individual components are user-defined, while in the NN-$\ell_1$-RJIVE the rank of each one of the individual components is automatically estimated via nuclear norm minimization. As opposed to the previous two models, the SRJIVE directly extracts the orthonormal bases of joint and individual components and improves their scalability. Finally, the RJIVE-M extends the SRJIVE in order to handle missing values.
- Based on the recovered joint and individual components from training data, two suitable optimization problems that extract the corresponding modes of variation (i.e., joint and individual components) of unseen test samples, are proposed in Section 4.
- To tackle the proposed optimization problems, algorithms based on the Alternating-Directions Method of Multipliers (ADMM) [17] are developed in Sections 3, 4, 5, and 6.
- We demonstrate the applicability of the proposed methods in three challenging computer vision tasks, namely facial expression synthesis, face age progression in 2D images and 3D data captured 'in-the-wild'. Experimental results corroborate the effectiveness of the proposed approach in Section 7.

*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 $\mathbf{x}_j$. Several norms and metrics will be used. The $\ell_1$- and the $\ell_2$-norms of $\mathbf{x}$ are defined as $\|\mathbf{x}\|_1 = \sum_i |x_i|$ and $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$, respectively. $|\cdot|$ denotes the absolute value operator. The matrix $\ell_1$ norm is defined as $\|\mathbf{X}\|_1 = \sum_i \sum_j |x_{ij}|$, 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 $\|\mathbf{X}\|_*$. The vector (matrix) $\ell_0$ -(quasi) norm returns the total number of non-zero elements in a vector (matrix). The rank function is denoted by rank($\cdot$).

The minimization of both the rank function and the $\ell_0$-norm are NP-hard problems [18], [19]. Consequently, the rank function and the $\ell_0$-norm are typically replaced by their convex surrogates [20], [21].

*Operators:* The solution of the several problems appeared in the paper relies on different (proximal) operators which are defined next. Let for any matrix $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the Singular Value Decomposition (SVD).

- Shrinkage operator [22]: $\mathcal{S}_\tau[\sigma] = \mathrm{sgn}(\sigma)\max(|\sigma| - \tau, 0)$.
- Singular Value Thresholding (SVT) operator [23]: $\mathcal{D}_\tau = \mathbf{U}\mathcal{S}_\tau\mathbf{V}^T$.
- Rank-r SVD operator: $\mathcal{Q}_r[\mathbf{X}] = \left[\mathbf{U}(:, 1:r)\mathbf{\Sigma}(1:r, 1:r)\mathbf{V}(:, 1:r)^T\right]$.
- Procrustes operator: $\mathcal{P}[\mathbf{D}] = \mathbf{G}\mathbf{R}^T$ (given the rank-r SVD of a matrix $\mathbf{D} = \mathbf{G}\mathbf{P}\mathbf{R}^T$).

## 2 Background

To make the paper self-contained, this section includes a brief review of the JIVE [12], COBE [13], and RCICA [14].

### 2.1 Joint and Individual Variation Explained (JIVE)

The JIVE recovers the joint and individual components among $M \geq 2$ datasets $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}, i = 1, 2, \ldots, M\}$, where $J$ is the number of samples of each dataset. In particular, each matrix is decomposed into two terms: a low-rank matrix $\mathbf{J}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}$ capturing *joint structure* among dataset and a low-rank matrix $\mathbf{A}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}$ capturing *individual structure* of each dataset. That is, $\mathbf{X}^{(i)} = \mathbf{J}^{(i)} + \mathbf{A}^{(i)}, i = 1, 2, \ldots, M$. Let $\mathbf{X}$ and $\mathbf{J}$ be $\sum_{i=1}^M d^{(i)} \times J$ matrices constructed by concatenation of the corresponding matrices, i.e., $\mathbf{X} = [\mathbf{X}^{(1)^T}, \mathbf{X}^{(2)^T}, \ldots, \mathbf{X}^{(M)^T}]^T$, $\mathbf{J} = [\mathbf{J}^{(1)^T}, \mathbf{J}^{(2)^T}, \ldots, \mathbf{J}^{(M)^T}]^T$, JIVE solves the rank-constrained least-squares problem [12]:

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

Problem (1) imposes rank constraints on the joint and individual components and requires the rows of $\mathbf{J}$ and $\{\mathbf{A}^{(i)}\}_{i=1}^M$ to be orthogonal. The intuition behind the orthogonality constraint stems for the fact that sample patterns responsible for joint structure between data types are unrelated to sample patterns responsible for individual structure [12]. By adopting the least squares error, the JIVE assumes Gaussian distributions with small variance [15]. Such an assumption rarely holds in real world data, where gross, non-Gaussian corruptions are in abundance. Consequently, the components obtained by employing the JIVE in the analysis of grossly corrupted data may be arbitrarily away from the true ones, thus degenerating their performance.

### 2.2 Common Orthogonal Basis Extraction (COBE)

A closely related method to the JIVE is the COBE which extracts the common and individual components of $M$ datasets of the same dimensions by solving a set of least-squares minimization problems [13]. More specifically, each dataset $\mathbf{X}^{(i)} \in \mathbb{R}^{J \times d^{(i)}}$ is factorized as $\mathbf{\Xi}^{(i)}\mathbf{\Lambda}^{(i)^T}$, where a column of $\mathbf{\Xi}^{(i)}$ signifies a latent variable to be found and $\mathbf{\Lambda}^{(i)}$ signifies a matrix of weights. $\mathbf{\Xi}^{(i)}$ is assumed to be decomposable in blocks as $\left[\bar{\mathbf{\Xi}}\tilde{\mathbf{\Xi}}^{(i)}\right]$ where

[1] A preliminary version of the present work has been proposed in [16], where the the main model and its algorithmic framework has been introduced. In this paper, we further investigate RJIVE and propose a unified model that directly extracts the orthonormal bases of joint and individual components and improves the scalability of the main model. Besides that we propose an extension of RJIVE for handling missing data. Moreover, new qualitative and quantitative experimental results are included in this paper.

$\bar{\Xi} \in \mathbb{R}^{n\times m}$, $\tilde{\Xi}^{(i)} \in \mathbb{R}^{n\times(d^{(i)}-m)}$ and $m \leq \min\{d^{(i)}, i = 1, \cdots, M\}$. In other words, $\bar{\Xi}$ is assumed to be common in all factorizations and hence it presents *joint structure*, while $\tilde{\Xi}^{(i)}$ is assumed to represent *individual structure*. Similarly, $\mathbf{\Lambda}^{(i)}$ splits into $\bar{\mathbf{\Lambda}}^{(i)}$ and $\tilde{\mathbf{\Lambda}}^{(i)}$. The optimization problem of the COBE takes the following form:

$$\begin{aligned} \min_{\bar{\Xi},\tilde{\Xi}^{(i)}} \quad & \sum_{i=1}^{M} \left\| \mathbf{X}^{(i)} - \bar{\Xi}\bar{\mathbf{\Lambda}}^{(i)^T} - \tilde{\Xi}^{(i)}\tilde{\mathbf{\Lambda}}^{(i)^T} \right\|_F^2 . \\ \text{s.t.} \quad & \bar{\Xi}^T\bar{\Xi} = \mathbf{I}, \{\tilde{\Xi}^{(i)^T}\tilde{\Xi}^{(i)} = \mathbf{I}, \bar{\Xi}^T\tilde{\Xi}^{(i)^T} = \mathbf{0}\}_{i=1}^{M}. \end{aligned} \tag{2}$$

Similarly to the JIVE, the utilization of the least square error renders the COBE non-robust against sparse, non-Gaussian errors.

### 2.3 Robust Correlated and Individual Component Analysis (RCICA)

The goal of the RCICA [14] is to extract both the *correlated* and the *individual* components between two known, high-dimensional datasets or views, namely $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)}\times J}\}_{i=1}^2$, in the presence of sparse noise (or errors). To this end, the RCICA seeks a decomposition of each data matrix $\{\mathbf{X}^{(i)}\}$ into three terms: $\mathbf{X}^{(i)} = \mathbf{C}^{(i)} + \mathbf{A}^{(i)} + \mathbf{E}^{(i)}, \quad i = 1, 2$. $\mathbf{C}^{(i)} \in \mathbb{R}^{d^{(i)}\times J}$ and $\mathbf{A}^{(i)} \in \mathbb{R}^{d^{(i)}\times J}$ are *low-rank* matrices, with rank($\mathbf{C}^{(i)}$) $\leq k_c$ and rank($\mathbf{A}^{(i)}$) $\leq k^{(i)}$ and mutually independent columns, capturing the correlated and individual components, respectively and $\mathbf{E}^{(i)} \in \mathbb{R}^{d^{(i)}\times J}$ is a *sparse* matrix accounting for the sparse noise.

To extract the correlated components $\mathbf{C}^{(i)} \in \mathbb{R}^{d^{(i)}\times J}$, the cost function of the Canonical Correlation Analysis (CCA) [7] is adopted. That is, by further decomposing the matrix $\{\mathbf{C}^{(i)}\}_{i=1}^2$ as: $\mathbf{C}^{(i)} = \mathbf{U}^{(i)} \ \mathbf{V}^{(i)^T} \ \mathbf{X}^{(i)}$, the maximally correlated components are derived by minimizing the CCA cost, namely $\frac{\lambda_c}{2}\|\mathbf{V}^{(1)^T}\mathbf{X}^{(1)} - \mathbf{V}^{(2)^T}\mathbf{X}^{(2)}\|_F^2$. Here, $\mathbf{U}^{(i)}$ are orthonormal bases, transforming the correlated components back to the observation space $\mathbf{X}^{(i)}$. Since the column space of the individual components $\mathbf{A}^{(i)}$ is desired to be orthogonal to the one of the correlated components, we have to enforce $\{\mathbf{Q}^{(i)^T}\mathbf{U}^{(i)}\}_{i=1}^2 = \mathbf{0}$, where $\mathbf{Q}^{(i)}$ are column orthonormal bases spanning the column space of the individual components $\mathbf{A}^{(i)}$. That is, $\mathbf{A}^{(i)} = \mathbf{Q}^{(i)} \ \mathbf{H}^{(i)}$.

Consequently, a natural estimator accounting for the upper-bounded rank of the correlated and independent components and the sparsity of $\{\mathbf{E}^{(i)}\}_{i=1}^2$ is to minimize the objective function of CCA, i.e., $\frac{1}{2}\|\mathbf{V}^{(1)^T}\mathbf{X}^{(1)} - \mathbf{V}^{(2)^T}\mathbf{X}^{(2)}\|_F^2$ as well as the rank of $\{\mathbf{C}^{(i)} = \mathbf{U}^{(i)} \ \mathbf{V}^{(i)^T} \ \mathbf{X}^{(i)}, \mathbf{A}^{(i)} = \mathbf{Q}^{(i)} \ \mathbf{H}^{(i)}\}_{i=1}^2$ and the number of non-zero entries of $\{\mathbf{E}^{(i)}\}_{i=1}^2$ measured by the $\ell_0$-(quasi) norm, e.g., [22]. To avoid the NP-hardness of rank and $\ell_0$-norm minimization, the nuclear- and the $\ell_1$-norms are typically adopted as surrogates to rank and $\ell_0$-norm, respectively [20], [21]. By employing the unitary invariance of the nuclear norm i.e., $\|\mathbf{Q}^{(i)}\mathbf{V}^{(i)^T}\|_* = \|\mathbf{V}^{(i)^T}\|_*$, the optimization problem of RCICA is formulated as the following constrained non-linear one:

$$\begin{aligned} \min_{\mathcal{V}} \quad & \sum_{i=1}^{2} \left[ \|\mathbf{V}^{(i)^T}\|_* + \lambda_*^{(i)}\|\mathbf{H}^{(i)}\|_* + \lambda_1^{(i)} \ \|\mathbf{E}^{(i)}\|_1 \right] \\ & + \frac{\lambda_c}{2}\|\mathbf{V}^{(1)^T}\mathbf{X}^{(1)} - \mathbf{V}^{(2)^T}\mathbf{X}^{(2)}\|_F^2, \\ \text{s.t.} \quad & (i) \ \ \mathbf{X}^{(i)} = \mathbf{U}^{(i)}\mathbf{V}^{(i)^T}\mathbf{X}^{(i)} + \mathbf{Q}^{(i)}\mathbf{H}^{(i)} + \mathbf{E}^{(i)}, \\ & (ii) \ \ \mathbf{V}^{(i)^T}\mathbf{X}^{(i)}\mathbf{X}^{(i)^T}\mathbf{V}^{(i)} = \mathbf{I}, \\ & (iii) \ \ \mathbf{U}^{(i)^T}\mathbf{U}^{(i)} = \mathbf{I}, \ \ \mathbf{Q}^{(i)^T}\mathbf{Q}^{(i)} = \mathbf{I}, \\ & (iv) \ \ \mathbf{Q}^{(i)^T}\mathbf{U}^{(i)} = \mathbf{0}, \ \ i = 1, 2, \end{aligned} \tag{3}$$

where the positive parameters $\lambda_c$, $\lambda_*^{(1)}$, $\lambda_*^{(2)}$, $\lambda_1^{(1)}$ and $\lambda_1^{(2)}$ control the correlation, rank and sparsity of the derived spaces and $\mathcal{V} = \{\mathbf{U}^{(i)}, \mathbf{V}^{(i)}, \mathbf{Q}^{(i)}, \mathbf{H}^{(i)}, \mathbf{E}^{(i)}\}_{i=1}^2$ collects the optimization variables. Constraints (ii) in (3) have been adopted from the CCA [7], while the constraints (iii) and (iv) ensure that both the recovered correlated and individual components are linearly independent.

Although the RCICA is robust to sparse, non-Gaussian error, its extension to more than two datasets is not trivial due to the orthogonality among the correlated and individual components and column orthonormality of the basis matrices $\mathbf{U}^{(i)}$ and $\mathbf{Q}^{(i)}$, $i = 1, 2, \ldots M$, with $M$ being the number of different views. This makes the resulting optimization problem highly-nonlinear and hence difficult to solve.

## 3 Robust JIVE

Consider data consisting of $M$ views, namely $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)}\times J}\}_{i=1}^M$, with $\mathbf{x}_j^{(i)} \in \mathbb{R}^{d^{(i)}}$, $j = 1, \ldots, J$ being a vectorized (visual) data sample, possibly contaminated by gross, sparse errors. The goal of the RJIVE is to robustly recover the joint components which are shared across all views as well as the components which are deemed individual for each view. That is:

$$\mathbf{X} = \mathbf{J} + \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T + \mathbf{E}, \tag{4}$$

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 (4), the joint and common components should be mutually incoherent, i.e., $\{\mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^M$. Assuming that the number of errors is bounded below some constant, the number of errors in the estimated components is similarly bounded and hence, a natural estimator accounting for the sparsity of the error matrix $\mathbf{E}$ is to minimize the number of the non-zero entries of $\mathbf{E}$ measured by the $\ell_0$-quasi norm [22]. However, as in case of the RCICA, to make the problem computationally tractable the $\ell_0$-norm is replaced by its convex surrogate, namely the $\ell_1$-norm. Therefore, the joint and individual components as well as the sparse error are recovered by solving the following constrained, non-linear optimization problem:

$$\begin{aligned} \min_{\mathbf{J},\{\mathbf{A}^{(i)}\}_{i=1}^M} \quad & \left\| \mathbf{X} - \mathbf{J} - \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T \right\|_1 . \\ \text{s.t.} \quad & \text{rank}(\mathbf{J}) = r, \{\text{rank}(\mathbf{A}^{(i)}) = r^{(i)}, \mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^M \end{aligned} \tag{5}$$

**Algorithm 1:** ADMM solver for (7) ($\ell_1$-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 $\rho$.
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
    $\mathbf{M} = \mathbf{X} - \left[\mathbf{A}_t^{(1)^T}, \cdots, \mathbf{A}_t^{(M)^T}\right]^T - \mathbf{E}_t + \mu_t^{-1}\mathbf{L}_t$;
    $\mathbf{J}_{t+1} = \mathcal{Q}_r[\mathbf{M}]$, $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \mathrm{svd}(\mathbf{M})$;
    $\mathbf{P} = \mathbf{I} - \mathbf{V}(:, 1:r)\mathbf{V}(:, 1:r)^T$;
    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
```

**Algorithm 2:** ADMM solver of (9) (NN-$\ell_1$-RJIVE).

```
Input      : Data $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$. Rank of joint
             component $r$. Parameter $\rho$.
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
    $\mathbf{J}_{t+1} = \mathcal{Q}_{\mathbf{r}}\left[\mathbf{X} - \left[\mathbf{A}_t^{(1)^T}, \cdots, \mathbf{A}_t^{(M)^T}\right]^T - \mathbf{E}_t + \frac{\mathbf{F}_t}{\mu_t}\right]$;
    for $i = 1 : M$ do
        $\mathbf{A}_{t+1}^{(i)} = \frac{\left(\mathbf{X}^{(i)} - \mathbf{J}_{t+1}^{(i)} - \mathbf{E}_t^{(i)} + \frac{\mathbf{F}^{(i)}}{\mu_t} + \mathbf{R}_t^{(i)} + \frac{\mathbf{Y}_t^{(i)}}{\mu_t}\right)\mathbf{P}}{2}$;
        $\mathbf{R}_{t+1}^{(i)} = \mathcal{D}_{1/\mu_t}\left[\mathbf{A}_{t+1}^{(i)} - \frac{\mathbf{Y}_t^{(i)}}{\mu_t}\right]$;
        $\mathbf{Y}_{t+1}^{(i)} = \mathbf{Y}_t^{(i)} + \mu_t(\mathbf{R}_{t+1}^{(i)} - \mathbf{A}_{t+1}^{(i)})$;
    end
    $\mathbf{E}_{t+1} =$
      $\mathcal{S}_{\frac{\lambda}{\mu_t}}\left[\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T + \frac{\mathbf{F}_t}{\mu_t}\right]$;
    $\mathbf{F}_{t+1} = \mathbf{F}_t + \mu_t(\mathbf{X} - \mathbf{J}_{t+1} - \left[\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}\right]^T -$
      $\mathbf{E}_{t+1}); t = t + 1$;
end
```

Clearly, (5) is a robust extension to JIVE [12], 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 since an extensive tuning procedure is required. To alleviate this issue, we propose a variant of (5), 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., $\{\mathrm{rank}(\mathbf{A}^{(i)}) \leq K^{(i)}\}_{i=1}^{M}$, problem (5) is relaxed to the following one:

$$\begin{aligned} \min_{\mathbf{J}, \{\mathbf{A}^{(i)}\}_{i=1}^{M}} \quad & \lambda \left\| \mathbf{X} - \mathbf{J} - \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T \right\|_1 + \\ & \sum_{i=1}^{M} \left\| \mathbf{A}^{(i)} \right\|_*, \text{s.t.} \quad \mathrm{rank}(\mathbf{J}) = r, \{\mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^{M}, \end{aligned} \tag{6}$$

where the rank function is replaced by its convex envelope, namely the nuclear norm and $\lambda > 0$ is a regularizer.

### 3.1 Optimization Algorithms

In this section, algorithms for solving (5) and (6) are developed.

To solve (5), the Alternating-Direction Method of Multipliers (ADMM) [17] is employed. To this end, problem (5) is reformulated to the following separable one:

$$\begin{aligned} \min_{\mathbf{J}, \{\mathbf{A}^{(i)}\}_{i=1}^{M}, \mathbf{E}} \quad & \|\mathbf{E}\|_1, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{J} + \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T + \mathbf{E}, \\ & \mathrm{rank}(\mathbf{J}) = r, \{\mathrm{rank}(\mathbf{A}^{(i)}) = r^{(i)}, \mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^{M}, \end{aligned} \tag{7}$$

where $\mathbf{E}$ is an auxiliary variable. To solve (7), the corresponding augmented Lagrangian function is given by:

$$\begin{aligned} & \mathcal{L}(\mathbf{J}, \{\mathbf{A}^{(i)}\}_{i=1}^{M}, \mathbf{E}, \mathbf{L}) = \|\mathbf{E}\|_1 - \frac{1}{2\mu}\|\mathbf{L}\|_F^2 \\ & + \frac{\mu}{2}\left\| \mathbf{X} - \mathbf{J} - \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T - \mathbf{E} + \frac{\mathbf{L}}{\mu} \right\|_F^2, \end{aligned} \tag{8}$$

where $\mathbf{L}$ is the Lagrange multipliers matrix related to the equality constraint in (7), and $\mu$ is a positive parameter. Subsequently, by employing the ADMM, (8) is minimized with respect to each variable in an alternating fashion and finally the Lagrange multipliers $\mathbf{L}$ are updated. The ADMM solver of (7) is outlined in Algorithm 1. Algorithm 1 terminates when $\|\mathbf{X} - \mathbf{J}_{t+1} - [\mathbf{A}_{t+1}^{(1)^T}, \cdots, \mathbf{A}_{t+1}^{(M)^T}]^T - \mathbf{E}_{t+1}\|_F^2 / \|\mathbf{X}\|_F^2$ is less than a predefined threshold $\epsilon$ or the number of iterations reach a maximum value.

To solve problem (6) via the ADMM, we firstly reformulate it as:

$$\begin{aligned} \min_{\mathbf{J}, \{\mathbf{A}^{(i)}, \mathbf{R}^{(i)}\}_{i=1}^{M}, \mathbf{E}} \quad & \sum_{i=1}^{M} \left\| \mathbf{R}^{(i)} \right\|_* + \lambda \|\mathbf{E}\|_1, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{J} + \left[\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}\right]^T + \mathbf{E}, \\ & \mathrm{rank}(\mathbf{J}) = r, \{\mathbf{R}^{(i)} = \mathbf{A}^{(i)}, \mathbf{J}\mathbf{A}^{(i)^T} = \mathbf{0}\}_{i=1}^{M} \end{aligned} \tag{9}$$

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. The ADMM solver of (9) is wrapped up in Algorithm 2 where $\mathbf{F}$, $\{\mathbf{Y}^{(i)}\}_{i=1}^{M}$ are the Lagrange multipliers related to the equality constraints in (9), and $\mu$ is a positive parameter. A convergence criterion similar to Algorithm 1 is employed. The augmented Lagrangian function of (9) as well as the derivation of the proposed Algorithm can be found in the supplementary material.

## 4 RJIVE-Based Reconstruction

Having recovered the individual and common components of the $M$ views or different datasets during training, we can exploit them in order to extract the joint and individual modes of variations of a test sample. For instance, the components recovered by applying the RJIVE on a set of facial images of $M$ different expressions can be utilized in order to reconstruct $M$ expressive images

**Algorithm 3:** ADMM-based solver of (10).

```
Input     : Input sample $\mathbf{t}$. Orthonormal bases
            $\mathbf{B}^{(i)} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{J}}^{(i)}}, \mathbf{D}^{(i)} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{A}}^{(i)}}$. Parameters
            $\lambda, \rho$.
Output    : Clean reconstructed image $\mathbf{y}$.
Initialize: Set $\{\mathbf{v}_0^{(n)}, \mathbf{c}_0^{(n)}\}_{n=1}^2$, $\{\mathbf{h}_0^{(n)}\}_{n=1}^4$, $\mathbf{y}_0$, and $\mathbf{e}_0$ to zero
            vectors, $t = 0$, $\mu_0 > 0$.
while not converged do
    for n=1:2 do
        $\mathbf{v}_{t+1}^{(n)} = \mathcal{S}_{\frac{1}{\mu_t}} \left[\mathbf{c}_t^{(n)} - \frac{\mathbf{h}_t^{(n)}}{\mu_t}\right]$;
    end
    $\tilde{\mathbf{t}}_1 = \mathbf{t} - \mathbf{D}^{(i)}\mathbf{c}_t^{(2)} - \mathbf{e}_t + \mathbf{h}_t^{(3)}\mu_t^{-1}$;
    $\tilde{\mathbf{t}}_2 = \mathbf{y} - \mathbf{D}^{(i)}\mathbf{c}_t^{(2)} + \mathbf{h}_t^{(4)}\mu_t^{-1}$;
    $\mathbf{c}_{t+1}^{(1)} = \frac{\mathbf{B}^{(i)^T}\left(\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2\right) + \mathbf{v}_{t+1}^{(1)} + \mathbf{h}_t^{(1)}\mu_t^{-1}}{3}$;
    $\tilde{\mathbf{t}}_1 = \mathbf{t} - \mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} - \mathbf{e}_t + \mathbf{h}_t^{(3)}\mu_t^{-1}$;
    $\tilde{\mathbf{t}}_2 = \mathbf{y} - \mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} + \mathbf{h}_t^{(4)}\mu_t^{-1}$;
    $\mathbf{c}_{t+1}^{(2)} = \frac{\mathbf{D}^{(i)^T}\left(\tilde{\mathbf{t}}_1 + \tilde{\mathbf{t}}_2\right) + \mathbf{v}_{t+1}^{(2)} + \mathbf{h}_t^{(2)}\mu_t^{-1}}{3}$;
    $\mathbf{y}_{t+1} = \max\left(\mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} + \mathbf{D}^{(i)}\mathbf{c}_{t+1}^{(2)} - \mathbf{h}_t^{(4)}/\mu_t, 0\right)$;
    $\mathbf{e}_{t+1} = \mathcal{S}_{\frac{\lambda}{\mu_t}} \left[\mathbf{t} - \mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} - \mathbf{D}^{(i)}\mathbf{c}_{t+1}^{(2)} + \mathbf{h}_t^{(3)}\mu_t^{-1}\right]$;
    $\mathbf{h}_{t+1}^{(1)} = \mathbf{h}_t^{(1)} + \mu_t(\mathbf{v}_{t+1}^{(1)} - \mathbf{c}_{t+1}^{(1)})$;
    $\mathbf{h}_{t+1}^{(2)} = \mathbf{h}_t^{(2)} + \mu_t(\mathbf{v}_{t+1}^{(2)} - \mathbf{c}_{t+1}^{(2)})$;
    $\mathbf{h}_{t+1}^{(3)} = \mathbf{h}_t^{(3)} + \mu_t(\mathbf{t} - \mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} - \mathbf{D}^{(i)}\mathbf{c}_{t+1}^{(2)} - \mathbf{e}_{t+1})$;
    $\mathbf{h}_{t+1}^{(4)} = \mathbf{h}_t^{(4)} + \mu_t(\mathbf{y} - \mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} - \mathbf{D}^{(i)}\mathbf{c}_{t+1}^{(2)})$;
    $\mu_{t+1} = \min(\mu_t\rho, 10^7)$;
end
```

$\{\mathbf{y}^{(i)}\}_{i=1}^M$ of an input face $\mathbf{t}$. The key motivation here is that the expression-related patterns of the image $\mathbf{t}$ in the expression $(i)$ lie in a linear subspace spanned by $\mathbf{D}^{(i)} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{A}}^{(i)}}$, where $\mathbf{D}^{(i)}$ has been obtained by applying the SVD onto the extracted $\mathbf{A}^{(i)}$ components. Therefore, the expression-related (individual) part of the test image $\mathbf{t}$ in expression $(i)$ can be represented as a linear combination of the orthonormal bases $\mathbf{D}^{(i)}$, i.e., $\mathbf{y}_{\text{individual}}^{(i)} \approx \mathbf{D}^{(i)}\mathbf{c}^{(2)}$ with $\mathbf{c}^{(2)} \in \mathbb{R}^{W_{\mathbf{A}}^{(i)} \times 1}$ being a sparse coefficient vector. Similarly, the joint part $\mathbf{y}_{\text{joint}}^{(i)}$ is expressed as a linear combination of the orthonormal bases $\mathbf{B}^{(i)} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{J}}^{(i)}}$, extracted from the corresponding joint component $\mathbf{J}^{(i)}$ i.e., $\mathbf{y}_{\text{joint}}^{(i)} \approx \mathbf{B}^{(i)}\mathbf{c}^{(1)}$, $\mathbf{c}^{(1)} \in \mathbb{R}^{W_{\mathbf{J}}^{(i)} \times 1}$. Thus, the expressive image $\mathbf{y}^{(i)}$ of the unseen input face $\mathbf{t}$ is reconstructed by solving the following constrained optimization problem:

$$\begin{aligned} \min_{\{\mathbf{c}^{(n)}, \mathbf{v}^{(n)}\}_{n=1}^2, \mathbf{y} \geq \mathbf{0}} \quad & \sum_{n=1}^{2} \left\|\mathbf{v}^{(n)}\right\|_1 + \lambda \left\|\mathbf{e}\right\|_1, \\ \text{s.t.} \quad & \{\mathbf{v}^{(n)} = \mathbf{c}^{(n)}\}_{n=1}^2 \\ \mathbf{t} = \mathbf{B}^{(i)}\mathbf{c}^{(1)} + \mathbf{D}^{(i)}\mathbf{c}^{(2)} + \mathbf{e}, \quad & \mathbf{y} = \mathbf{B}^{(i)}\mathbf{c}^{(1)} + \mathbf{D}^{(i)}\mathbf{c}^{(2)} \end{aligned} \tag{10}$$

where $\lambda$ is a positive parameter that balances the norms, $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ are auxiliary variables which are employed in order to make the problem separable, $\mathbf{y}$ corresponds to the non-negative clean reconstruction, and $\mathbf{e}$ is an error term accounting for the gross, non-Gaussian sparse noise. Equation (10) resembles the dense error correction model proposed in [24], which is suitable for guaranteed recovery of sparse representations from high-dimensional measurements, such as images of high resolution (e.g., 22000 pixels in this paper) in the presence of noise. The ADMM solver of (10) is outlined in Algorithm 3. Algorithm 3 terminates when $\left\|\mathbf{t} - \mathbf{B}^{(i)}\mathbf{c}_{t+1}^{(1)} - \mathbf{D}^{(i)}\mathbf{c}_{t+1}^{(2)} - \mathbf{e}_{t+1}\right\|_2^2 / \|\mathbf{t}\|_2^2$ is less than a predefined threshold $\epsilon$ or the number of iterations reached. The augmented Lagrangian function of (10) can be found in the supplementary material.

## 5 Scalable RJIVE

The computational complexity of the vanilla JIVE as well as the $\ell_1$-RJIVE and NN-$\ell_1$-RJIVE at each iteration is $\mathcal{O}(\max(q^2 J, qJ^2)) + \sum_{i=1}^M \mathcal{O}(\max(d^{(i)^2} J, d^{(i)} J^2)) = \mathcal{O}(\max(q^2 J, qJ^2))$ due to SVD. Clearly, this is computationally prohibitive when dimensionality of images $\{d^{(i)}\}_{i=1}^M$ becomes very large, e.g., 22500 in our case. To alleviate the aforementioned computational complexity issue and at the same time learn the orthonormal bases that are used for reconstruction, we propose to factorize matrices $\mathbf{J}, \{\mathbf{A}^{(i)}\}_{i=1}^M$ as products of orthonormal bases matrices $\mathbf{B} \in \mathbb{R}^{(d^{(1)} + \cdots d^{(M)}) \times W_{\mathbf{J}}}, \mathbf{B}^T\mathbf{B} = \mathbf{I}$, $\{\mathbf{D}^{(i)} \in \mathbb{R}^{d^{(i)} \times W_{\mathbf{A}}^{(i)}} \mathbf{D}^{(i)^T}\mathbf{D}^{(i)} = \mathbf{I}\}_{i=1}^M$ and low-rank coefficients matrices $\mathbf{G} \in \mathbb{R}^{W_{\mathbf{J}} \times J}, \{\mathbf{C}^{(i)} \in \mathbb{R}^{W_{\mathbf{A}}^{(i)} \times J}\}_{i=1}^M$ such that $\mathbf{J} = \mathbf{B}\mathbf{G}$ and $\{\mathbf{A}^{(i)} = \mathbf{D}^{(i)}\mathbf{C}^{(i)}\}_{i=1}^M$. It can be easily shown that the constraints are now written as $\{\mathbf{J}\mathbf{A}^{(i)^T}\}_{i=1}^M = \mathbf{G}\mathbf{C}^{(i)^T} = \mathbf{0}$ and $\text{rank}(\mathbf{J}) = \text{rank}(\mathbf{B}\mathbf{G}) = \text{rank}(\mathbf{G}) = r$. In addition, due to the unitary invariance property of the nuclear norm we have $\|\mathbf{A}^{(i)}\|_* = \|\mathbf{D}^{(i)}\mathbf{C}^{(i)}\|_* = \|\mathbf{C}^{(i)}\|_*$. Therefore, by incorporating the factorizations of joint and individual components, the optimization problem (9) now reformulates as follows:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{G}, \{\mathbf{D}^{(i)}, \mathbf{C}^{(i)}, \mathbf{\Delta}^{(i)}\}_{i=1}^M, \mathbf{E}} \quad & \sum_{i=1}^{M} \left\|\mathbf{\Delta}^{(i)}\right\|_* + \lambda \left\|\mathbf{E}\right\|_1, \\ \text{s.t.} \quad \mathbf{X} = \mathbf{B}\mathbf{G} + & \left[\left(\mathbf{D}^{(1)}\mathbf{C}^{(1)}\right)^T \cdots, \left(\mathbf{D}^{(M)}\mathbf{C}^{(M)}\right)^T\right]^T + \mathbf{E}, \\ & \text{rank}(\mathbf{G}) = r, \mathbf{B}^T\mathbf{B} = \mathbf{I}, \\ \{\mathbf{\Delta}^{(i)} = \mathbf{C}^{(i)}, & \mathbf{G}\mathbf{C}^{(i)^T} = \mathbf{0}, \mathbf{D}^{(i)^T}\mathbf{D}^{(i)} = \mathbf{I}\}_{i=1}^M, \end{aligned} \tag{11}$$

where $\{\mathbf{\Delta}^{(i)} \in \mathbb{R}^{W_{\mathbf{A}}^{(i)} \times J}\}_{i=1}^M$ and $\{\mathbf{\Delta}^{(i)} = \mathbf{C}^{(i)}\}_{i=1}^M$, are auxiliary variables and the corresponding constraints, respectively.

The ADMM solver of the proposed SRJIVE method is outlined in Algorithm 4, where $\mathbf{\Gamma}$ and $\{\mathbf{Z}^{(i)}\}_{i=1}^M$ are the Lagrangian multipliers related to the equality constraints of (11) (the Lagrange function corresponding to problem (11) can be found in the supplementary material).

The computational complexity of Algorithm 4 is dominated by the cost of the SVD involved in the computation of SVT and Procrustes operators in Steps 4 and 5, respectively. Therefore, the computational complexity of each iteration is $\mathcal{O}(\max(W_{\mathbf{J}}^2 J, W_{\mathbf{J}} J^2))$ and $\mathcal{O}(\max(q^2 W_{\mathbf{J}}, q W_{\mathbf{J}}^2))$, respectively. Given that $W_{\mathbf{J}} \ll q = d^{(1)} + \cdots d^{(M)}$ (in this paper $q = 225000$ and $W_{\mathbf{J}} \leq 600$), which implies $W_{\mathbf{J}} J + q W_{\mathbf{J}} \ll qJ$, the proposed scalable version of RJIVE, i.e., the SRJIVE, has a significantly reduced computational cost compared to that of JIVE and RJIVE.

Regarding the convergence of the presented Algorithms 2, 1, 4, there is currently no theoretical proof known for the ADMM in problems with more than two blocks of variables. However, ADMM has been applied successfully in non-linear optimization problems in practice [14], [25], [26], [27], [28]. In addition, the thorough experimental evaluation of the proposed methods presented in Section 7, indicates that the obtained solutions are good for the data upon which RJIVE was tested.

**Algorithm 4:** ADMM solver of (11) (Scalable NN-$\ell_1$-RJIVE, SRJIVE).

```
Input     : Data $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$. Rank of joint component $r$. Number of bases to be extracted from the Joint and Individual
            components $W_{\mathbf{J}}$ and $W_{\mathbf{A}}^{(i)}$, respectively. Parameter $\rho$.
Output    : Orthonormal Joint and Individual bases matrices $\mathbf{B}$, $\{\mathbf{D}^{(i)}\}_{i=1}^{M}$. Coefficient matrices $\mathbf{G}$, $\{\mathbf{C}^{(i)}\}_{i=1}^{M}$.
Initialize: Set $\mathbf{G}_0$, $\mathbf{B}_0$, $\{\mathbf{\Delta}_0^{(i)}, \mathbf{D}_0^{(i)}, \mathbf{C}_0^{(i)}, \mathbf{Z}_0^{(i)}\}_{i=1}^{M}$, $\mathbf{E}_0$, $\mathbf{\Gamma}_0$ 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
    $\mathbf{M} = \mathbf{B}_t^T \left(\mathbf{X} - \left[\left(\mathbf{D}_t^{(1)}\mathbf{C}_t^{(1)}\right)^T \cdots, \left(\mathbf{D}_t^{(M)}\mathbf{C}_t^{(M)}\right)^T\right]^T - \mathbf{E}_t + \mu_t^{-1}\mathbf{\Gamma}_t\right); [\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \mathrm{svd}(\mathbf{M});$
    $\mathbf{G}_{t+1} = \mathcal{Q}_r[\mathbf{M}];$
    $\mathbf{B}_{t+1} = \mathcal{P}\left[\left(\mathbf{X} - \left[\left(\mathbf{D}_t^{(1)}\mathbf{C}_t^{(1)}\right)^T \cdots, \left(\mathbf{D}_t^{(M)}\mathbf{C}_t^{(M)}\right)^T\right]^T - \mathbf{E}_t + \mu_t^{-1}\mathbf{\Gamma}_t\right)\mathbf{G}_{t+1}^T\right];$
    $\mathbf{M} = \mathbf{X} - \mathbf{B}_{t+1}\mathbf{G}_{t+1} - \mathbf{E}_t + \mu_t^{-1}\mathbf{\Gamma}_t;$
    for n=1:M do
        $\mathbf{D}_{t+1}^{(i)} = \mathcal{P}\left[\mathbf{M}^{(i)}\mathbf{C}_t^{(i)^T}\right]; \mathbf{C}_{t+1}^{(i)} = 0.5\left(\mathbf{D}_{t+1}^{(i)^T}\mathbf{M}^{(i)} + \mathbf{\Delta}_t^{(i)} + \mu_t^{-1}\mathbf{Z}_t^{(i)}\right)\left(\mathbf{I} - \mathbf{V}\mathbf{V}^T\right);$
        $\mathbf{\Delta}_{t+1}^{(i)} = \mathcal{D}_{\frac{1}{\mu_t}}\left[\mathbf{C}_{t+1}^{(i)} - \mu^{-1}\mathbf{Z}_t^{(i)}\right];$
        $\mathbf{Z}_{t+1}^{(i)} = \mathbf{Z}_{t+1}^{(i)} + \mu_t\left(\mathbf{\Delta}_{t+1}^{(i)} - \mathbf{C}_{t+1}^{(i)}\right);$
    end
    $\mathbf{E} = \mathcal{S}_{\frac{\lambda}{\mu_t}}\left[\mathbf{X} - \mathbf{B}_{t+1}\mathbf{G}_{t+1} - \left[\left(\mathbf{D}_{t+1}^{(1)}\mathbf{C}_{t+1}^{(1)}\right)^T \cdots, \left(\mathbf{D}_{t+1}^{(M)}\mathbf{C}_{t+1}^{(M)}\right)^T\right]^T + \mu_t^{-1}\mathbf{\Gamma}_t\right];$
    $\mathbf{\Gamma}_{t+1} = \mathbf{\Gamma}_t + \mu_t\left(\mathbf{X} - \mathbf{B}_{t+1}\mathbf{G}_{t+1} - \left[\left(\mathbf{D}_{t+1}^{(1)}\mathbf{C}_{t+1}^{(1)}\right)^T \cdots, \left(\mathbf{D}_{t+1}^{(M)}\mathbf{C}_{t+1}^{(M)}\right)^T\right]^T - \mathbf{E}_{t+1}\right);$
    $\mu_{t+1} = \min(\rho \cdot \mu_t, 10^7); t = t + 1;$
end
```

## 6 RJIVE with missing values and application to face aging using 3D Morphable Models

3D Morphable Models (3DMMs) are statistical deformable models of the 3D shape and appearance of the human face [29]. Typically, a 3DMM consists of PCA models for shape and appearance, as well as a camera projection model. More specifically, the shape model describes facial meshes that consist of $L$ vertexes and is built by applying dense registration on a set of training meshes followed by PCA [29]. An instance of the shape model can be expressed as the linear combination of a mean shape $\bar{\mathbf{s}}$ and the subspace $\mathbf{U}_s$ with parameters $\mathbf{p}$ as $\mathbf{s} = \bar{\mathbf{s}} + \mathbf{U}_s\mathbf{p}$. Similarly, the texture model is a linear PCA model that describes the texture associated with the shape model and can be constructed from captured 3D texture as in [29], or from single 2D images as in [30]. Moreover, the camera model maps a 3D mesh on the image plane, utilizing an orthographic or a perspective transformation $W(\mathbf{p}, \mathbf{c})$, where $\mathbf{c}$ are the camera parameters. Fitting a 3DMM into a new image is an iterative process, where the model parameters (regarding shape, texture, and camera) are updated at each iteration. Typically, the fitting procedure is formulated as a Gauss-Newton optimization problem, where the main task is the minimization of the error between the input and the reconstructed image [30].

The extraction of 3D texture from single images commences with fitting a 3DMM on them. Then a UV texture map is calculated by projecting the reconstructed 3D shape on the image plane and subsequently sampling the image at the locations of the shape's vertexes. However, extracting the 3D texture from a 2D image in this way leads to incomplete 3D texture representations, mainly due to the presence of self-occlusions, especially when the person depicted in the image is not in a frontal pose. Therefore, data collected with the aforementioned technique include missing values. In order to specify the location (i.e., image coordinates) of the missing values in a UV texture image, a self-occlusion mask for each image is calculated by casting a ray from the camera to each vertex of the reconstructed shape. Each element of the extracted mask denotes whether a value of the UV texture map is missing or not (please see the first column of Figure 11 for a visualization of the extracted UV space).

Even though RJIVE can robustly recover joint and individual components in the presence of sparse non-Gaussian errors of large magnitude, it is not able to handle data with missing values. To overcome this limitation of the RJIVE, we propose the RJIVE-Missing (RJIVE-M). Consider $M$ datasets of different ages $\{\mathbf{X}^{(i)} \in \mathbb{R}^{d^{(i)} \times J}\}_{i=1}^{M}$, with $\mathbf{x}_j^{(i)} \in \mathbb{R}^{d^{(i)}}$, being a vectorized form of the $j$-th gross corrupted and incomplete UV texture, $j = 1, \ldots, J$, that displays a face within the $i$-th age group, $i = 1, \ldots M$. The goal of the RJIVE-M is not only to recover the joint and individual components but also to perform completion on the UV textures with missing values. To this end, problem (11) is reformulated to the following one:

$$
\begin{aligned}
&\min_{\mathbf{B}, \mathbf{G}, \{\mathbf{D}^{(i)}, \mathbf{C}^{(i)}, \mathbf{\Delta}^{(i)}\}_{i=1}^{M}, \mathbf{E}} \quad \sum_{i=1}^{M} \left\|\mathbf{\Delta}^{(i)}\right\|_* + \lambda \left\|\mathbf{W} \circ \mathbf{E}\right\|_1, \\
&\text{s.t.} \quad \mathbf{X} = \mathbf{B}\mathbf{G} + \left[\left(\mathbf{D}^{(1)}\mathbf{C}^{(1)}\right)^T \cdots, \left(\mathbf{D}^{(M)}\mathbf{C}^{(M)}\right)^T\right]^T + \mathbf{E}, \\
&\qquad \mathrm{rank}(\mathbf{G}) = r, \mathbf{B}^T\mathbf{B} = \mathbf{I}, \\
&\qquad \{\mathbf{\Delta}^{(i)} = \mathbf{C}^{(i)}, \mathbf{G}\mathbf{C}^{(i)^T} = \mathbf{0}, \mathbf{D}^{(i)^T}\mathbf{D}^{(i)} = \mathbf{I}\}_{i=1}^{M},
\end{aligned}
\tag{12}
$$

where $\circ$ denotes the Hadamard (element-wise) product and $\mathbf{W} = \left[\mathbf{W}^{(1)^T}, \cdots, \mathbf{W}^{(M)^T}\right]^T \in \mathbb{R}^{q \times J}$, $\mathbf{W}^{(i)} = [\mathbf{w}_1^{(i)}, \mathbf{w}_2^{(i)}, \cdots, \mathbf{w}_J^{(i)}] \in \{0, 1\}^{q \times J}$, with $\mathbf{w}_j^{(i)}$ being a vectorized form of the self-occlusion mask that corresponds to the $j$-th UV texture of the $i$-th dataset. The Algorithm for solving the proposed RJIVE-M problem is similar to the SRJIVE one and has the same complexity and convergence criterion. The only difference is in the updating step of the error matrix $\mathbf{E}$. More specifically, the following additional step is performed after executing the step 13 of Algorithm 4: $\mathbf{E} = \mathbf{W} \circ \mathbf{E} + \overline{\mathbf{W}} \circ [\mathbf{X} - \mathbf{B}_{t+1}\mathbf{G}_{t+1} - [(\mathbf{D}^{(1)}_{t+1}\mathbf{C}^{(1)}_{t+1})^T \cdots, (\mathbf{D}^{(M)}_{t+1}\mathbf{C}^{(M)}_{t+1})^T]^T + \mu_t^{-1}\mathbf{\Gamma}_t]$.

Similarly, the presented RJIVE-based reconstruction method can be also extended to handle missing values in a test image. To this end, given a test sample with missing values (e.g., facial UV texture) and the vectorized form of the corresponding occlusion mask $\mathbf{w}$, problem (10) is extended to the following one:

$$\begin{aligned} \min_{\{\mathbf{c}^{(n)}, \mathbf{v}^{(n)}\}_{n=1}^{2}, \mathbf{y} \geq \mathbf{0}} \quad & \sum_{n=1}^{2} \left\| \mathbf{v}^{(n)} \right\|_1 + \lambda \left\| \mathbf{w} \circ \mathbf{e} \right\|_1 . \\ \text{s.t.} \quad & \{\mathbf{v}^{(n)} = \mathbf{c}^{(n)}\}_{n=1}^{2} \\ \mathbf{t} = \mathbf{B}^{(i)}\mathbf{c}^{(1)} + \mathbf{D}^{(i)}\mathbf{c}^{(2)} + \mathbf{e}, & \quad \mathbf{y} = \mathbf{B}^{(i)}\mathbf{c}^{(1)} + \mathbf{D}^{(i)}\mathbf{c}^{(2)} \end{aligned} \tag{13}$$

An ADMM-based solver similar to the Algorithm 3 is employed in order to solve problem (13). More specifically, the update step of the error vector performed in step 12 of the Algorithm 3 is followed by $\mathbf{e}_{t+1} = \mathbf{w} \circ \mathbf{e} + \overline{\mathbf{w}} \circ \left[\mathbf{t} - \mathbf{B}^{(i)}\mathbf{c}^{(1)}_{t+1} - \mathbf{D}^{(i)}\mathbf{c}^{(2)}_{t+1} + \mathbf{h}^{(3)}_t \mu_t^{-1}\right]$.

# 7 Experimental Evaluation

The performance of the proposed RJIVE method is assessed on synthetic data corrupted by both Gaussian and sparse, non-Gaussian noise (Section 7.1), as well as on data captured under constrained and ‘in-the-wild’ conditions with applications to (a) *facial expression synthesis*, (b) 2D and (c) 3D *face age progression*.

TABLE 1: Parameters used in the conducted experiments.

| Section | $r$ | $\lambda$ | $W^{(i)}_{\mathbf{J}}$ | $W^{(i)}_{\mathbf{A}}$ | $\lambda$ | $\epsilon$ |
|---|---|---|---|---|---|---|
| 7.2.1 | 20 | $\frac{1}{\sqrt{\max(q,J)}}$ | 70 | 70 | 0.03 | $10^{-5}$ |
| 7.2.2 | 150 | | 300 | 300 | | |
| 7.3 | 300 | | 600 | 600 | | |

## 7.1 Synthetic

In this section, the ability of RJIVE to robustly recover the common 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 $\ell_1$-RJIVE and Nuclear-Norm regularized (NN-$\ell_1$-RJIVE) for a varying number of dimensions, joint and individual ranks, are reported in Table 2. The corresponding RRE obtained by JIVE [12], [31], COBE [13], and RCICA [14] 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-$\ell_1$-RJIVE successfully recovered all components by utilizing only the true rank of the joint component. In the other hand, all the other methods require knowledge regarding the true rank for both joint and individual components. Furthermore, the SRJIVE achieved same results to the NN-$\ell_1$-RJIVE by reducing the computation time more than five times. Based on the performance of SRJIVE on the synthetic data, we decided to utilize it in the experiments described bellow and refer it as RJIVE hereafter.

Furthermore, we tested the RJIVE on synthetic data contaminated by Gaussian error. The RJIVE can implicitly handle data contaminated by Gaussian noise by vanishing the error term. That is we set the regularizer $\lambda$ in problems (7), (9), (11) $\lambda \rightarrow \infty$ i.e. $\mathbf{E} = 0$. In such case, the Frobenius norm corresponding to the equality constraints $\mathbf{X} = \mathbf{J} + [\mathbf{A}^{(1)^T}, \cdots, \mathbf{A}^{(M)^T}]^T + \mathbf{E}$, $\mathbf{X} = \mathbf{B}\mathbf{G} + [(\mathbf{D}^{(1)}\mathbf{C}^{(1)})^T, \cdots, \mathbf{A}^{(M)^T}]^T + \mathbf{E}$ appearing in the corresponding augmented Lagrangian functions are deemed as the appropriate regularizer for handling Gaussian noise. The RRE of all compared methods are reported in Table 2. As it can be seen, the proposed methods accurately recovered both the joint and individual components.

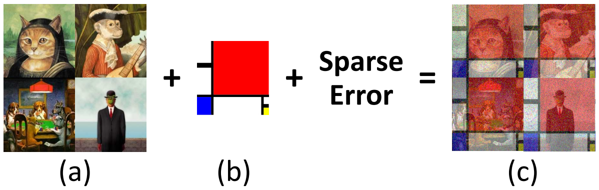


Fig. 1: Procedure followed to generate data contaminated by sparse, non-Gaussian noise.

The efficiency of the JIVE and RJIVE methods was qualitatively evaluated on real data contaminated by sparse, non-Gaussian noise. In order to generate the corrupted data, we firstly superimposed the paintings of Figure 1(a) with the painting appeared in Figure 1(b) and subsequently a sparse error matrix was added. In each image the error matrix has $20\%$ non-zero entries being sampled independently from $\mathcal{N}(0, 1)$. Then, the concatenation of the generated paintings (Figure 1(c)) was given as input to the JIVE and RJIVE. The joint and individual components as well as the corresponding error matrices obtained from the compared methods are depicted in Figure 2[2]. As it can be observed, the RJIVE accurately recovered both the joint and individual components. On the other hand, the joint components extracted from JIVE are not accurate, while the corresponding individual ones are contaminated by the sparse error. This is due to the fact that the JIVE is not robust to sparse, non-Gaussian noise.

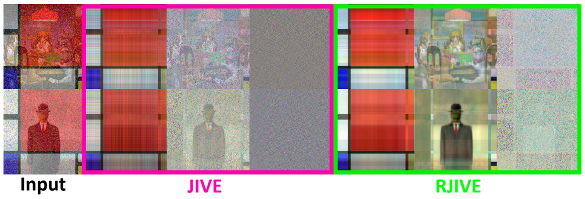


Fig. 2: Joint, individual components, and error matrices produced by the compared JIVE and RJIVE methods.

[2] Additional results can be found in the supplementary material.

TABLE 2: Quantitative recovering results produced by JIVE [12], COBE [13], RCICA [14], $\ell_1$-RJIVE (7), and NN-$\ell_1$-RJIVE (9) under Gaussian and gross non-Gaussian noise. Each compared method was applied on the same data generated by utilizing each set of parameters. The average recovery accuracy and computation time (in CPU seconds) were computed by repeating the experiment $10$ times.

| $\left(d^{(1)}, d^{(2)}, J, r, r^{(1)}, r^{(2)}\right)$ | **Method** | $\left\{ \left\Vert \mathbf{J}_*^{(i)} - \mathbf{J}^{(i)} \right\Vert_F^2 / \left\Vert \mathbf{J}_*^{(i)} \right\Vert_F^2 \right\}_{i=1}^2$ | | $\left\{ \left\Vert \mathbf{A}_*^{(i)} - \mathbf{A}^{(i)} \right\Vert_F^2 / \left\Vert \mathbf{A}_*^{(i)} \right\Vert_F^2 \right\}_{i=1}^2$ | | **Time** (in CPU seconds) | |
|---|---|---|---|---|---|---|---|
| | | **non-Gaussian** | **Gaussian** | **non-Gaussian** | **Gaussian** | **non-Gaussian** | **Gaussian** |
| $(500, 500, 500, 5, 10, 10)$ | **COBE** | $3.6403$ | $1.0927$ | $1.0975$ | $1.0002$ | 0.06 | 0.07 |
| | **JIVE** | $0.5424$ | $1.3558e-04$ | $0.9349$ | $2.0782e-04$ | 4.62 | 1.22 |
| | **RCICA** | $-$ | $-$ | $7.1379e-07$ | $5.6337e-03$ | 1.14 | 1.36 |
| | **$\ell_1$-RJIVE** | $5.5628e-08$ | $1.3558e-04$ | $3.5073e-08$ | $2.0782e-04$ | 3.11 | 4.78 |
| | **NN-$\ell_1$-RJIVE** | $5.1515e-08$ | $1.4720e-04$ | $4.3416e-08$ | $3.3904e-04$ | 4.06 | 5.06 |
| | **SRJIVE** | $2.7770e-08$ | $1.6564e-04$ | $3.8706e-08$ | $2.0012e-04$ | 0.91 | 1.97 |
| $(1000, 1000, 1000, 10, 20, 20)$ | **COBE** | $4.9982$ | $1.08555$ | $1.1890$ | $0.9982$ | 0.122 | 0.11 |
| | **JIVE** | $0.8398$ | $1.8880e-04$ | $1.4810$ | $2.9261e-04$ | 14.69 | 4.45 |
| | **RCICA** | $-$ | $-$ | $6.7260e-07$ | $9.6371e-04$ | 6.36 | 5.84 |
| | **$\ell_1$-RJIVE** | $8.5033e-08$ | $1.8879e-04$ | $5.5423e-08$ | $2.9260e-04$ | 8.23 | 18.01 |
| | **NN-$\ell_1$-RJIVE** | $9.3804e-08$ | $2.0738e-04$ | $7.6262e-08$ | $1.1801e-04$ | 17.34 | 23.11 |
| | **SRJIVE** | $6.8905e-08$ | $2.3406e-04$ | $6.0017e-08$ | $1.2041e-04$ | 3.99 | 9.05 |
| $(2000, 2000, 2000, 20, 40, 40)$ | **COBE** | $6.9981$ | $1.088417$ | $1.3469$ | $0.9976$ | 0.83 | 0.69 |
| | **JIVE** | $1.3961$ | $2.6525e-04$ | $2.1977$ | $4.1133e-04$ | 203.25 | 49.06 |
| | **RCICA** | $-$ | $-$ | $5.9359e-05$ | $7.6497e-03$ | 48.51 | 49.86 |
| | **$\ell_1$-RJIVE** | $1.2305e-07$ | $2.6525e-04$ | $1.0512e-07$ | $4.1133e-04$ | 142.44 | 160.21 |
| | **NN-$\ell_1$-RJIVE** | $8.8570e-08$ | $2.9010e-04$ | $9.1058e-08$ | $5.6000e-04$ | 110.36 | 120.01 |
| | **SRJIVE** | $9.7434e-08$ | $2.7074e-04$ | $1.0117e-07$ | $5.1173e-04$ | 18.96 | 43.07 |

## 7.2 Facial Expression Synthesis

In this section, we investigate the ability of the RJIVE to synthesize a set of different expressions of a given facial image. Consider $M$ datasets, 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 dataset should be put in correspondence. Therefore, their $N = 68$ facial landmark points are localized using the detector [32], [33] and subsequently employed to compute a mean reference shape. Then, the faces of each dataset are warped into a corresponding reference shape by using the piecewise affine warp function $\mathcal{W}(\cdot)$ [34]. After applying the RJIVE on the warped datasets, the recovered components can be used to synthesize $M$ different expressions of an unseen subject. To do that, the new (unseen) facial image is warped to the reference frame that corresponds to the expression that we want to synthesize and subsequently is given as input to solve (10).

The performance of RJIVE in FES task is assessed by conducting inner- and cross-databases experiments on MPIE [35], CK+ [36], and 'in-the-wild' facial images collected from the internet (ITW). The synthesized expressions obtained by RJIVE are compared to those obtained by the state-of-the-art BKRRR [37] 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 learned 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.

### 7.2.1 Controlled Conditions

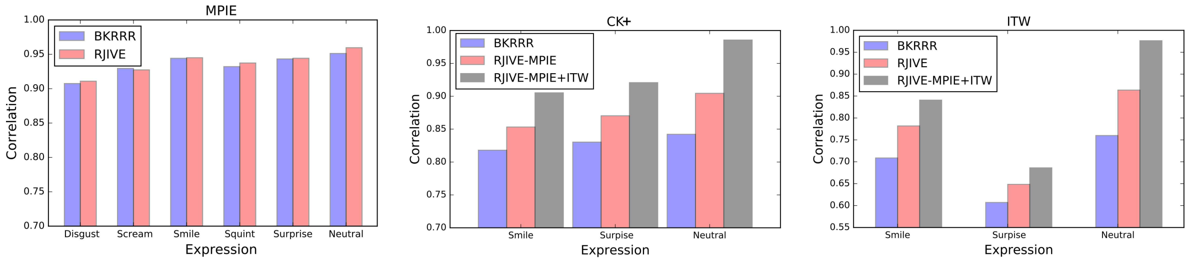
In the first experiment, $534$ frontal images of MPIE database that depict $89$ subjects under six expressions (i.e., 'Neutral', 'Scream', 'Squint', 'Surprise', 'Smile', 'Disgust') were employed to train both RJIVE and BKRRR. Then, all expressions of $58$ unseen subjects from the same database were synthesized by using their images corresponding to 'Neutral' expressions. In Figure 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, the performance of both methods is compared by performing a cross-database experiment on the CK+ database. More specifically, we employed the 'Neutral', 'Smile', and 'Surprise' images of MPIE for training purposes while images of $69$ subjects (three images per subject) of CK+ were used as test ones. In Figure 3(b) we can see that RJIVE outperforms BKRRR by a large margin. This is due to the fact that the BKRRR performs the regression based on how close is the unseen 'Neutral' face to the training ones. Therefore, in cases where the unseen subjects (e.g., subjects of CK+) present enough differences compared to the training ones (e.g., subjects of MPIE), the synthesized expressions are characterized as non-accurate. The synthesized expressions of subjects '$014$' and '$015$' from MPIE produced by the BKRRR and RJIVE are visualized in Figure 4. Clearly, the proposed method produces expressed images of higher quality compared to the BKRRR.

Finally, the recovering accuracy of JIVE and RJIVE in FES was also qualitatively assessed. To this end, the images used in the previous experiments were contaminated by sparse error were subsequently provided to the compared methods. Figure 5[2] displays the obtained components and the corresponding error matrices. Clearly, the proposed RJIVE method successfully recovered all the components. It is worth mentioning that the RJIVE removed the added sparse noise as well as the occlusions produced by eyeglasses and paintings (please see red dotted boxes). Instead, the JIVE was not able to remove neither the added noise nor the occlusions.

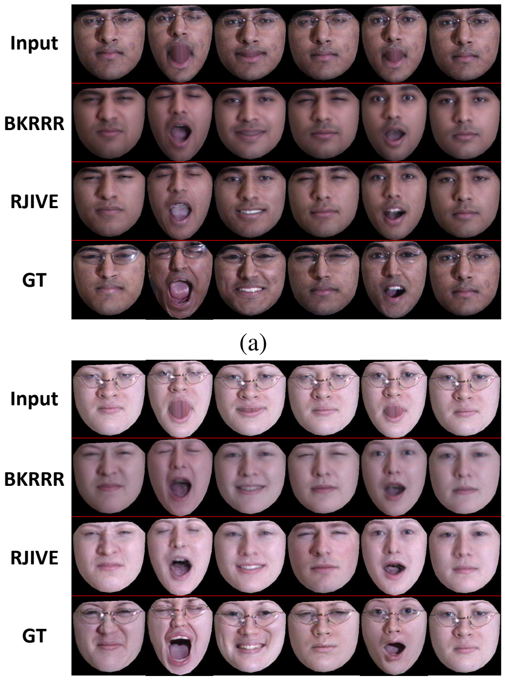
### 7.2.2 'In-The-Wild' Conditions

As an additional experiment, 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 methods trained on MPIE. Figure 3(c) depicts the obtained correlations for each subject. Clearly, the RJIVE outperforms the BKRRR. Compared to the previous experiments, there is a drop in the performance for both methods.



(a) (b) (c)

Fig. 3: Mean average correlation achieved by JIVE and BKRRR methods on (a) MPIE, (b) CK+, and (c) ITW databases.



(a)

(b)

Fig. 4: Synthesized expressions of MPIE's subject (a) '014' and (b) '015' produced by the BKRRR and RJIVE methods.

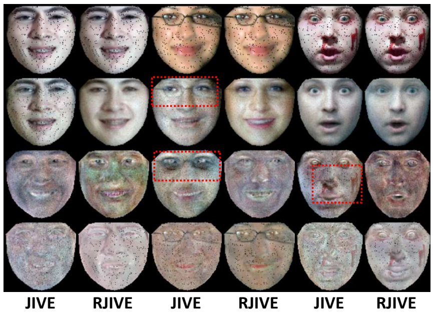


Fig. 5: Joint, individual components and error matrices produced by the compared JIVE and RJIVE methods.

This is attributed to the fact that the methods were trained by employing only images captured under controlled conditions. 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 on '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 the WWB database [38] (400 images for each expression). As it can be observed in Figure 3(c), the 'in-the-wild' train set improved the accuracy of RJIVE in both CK+ and ITW datasets. Figure 6 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 subsequently, the synthesized expressions were warped and fused with the actual images [39]. Clearly, the produced expressions are characterized by high quality of both expression and identity information. It is worth mentioning that RJIVE synthesizes almost perfectly the input images without using any kind of information about the depicted subject.

## 7.3 Face Age Progression 'In-The-Wild'

### 7.3.1 2D age progression of an unseen subject

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 [40], [41]. 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). A new database that overcomes the aforementioned problems was recently proposed in [42]. The AgeDB was manually collected and annotated. It consists of 16.488 images that depict 568 subjects from 0 to 101 years old. Annotations in terms of age and identity of the depicted subjects are provided. On average, there are 29 images that span 50.3 years for each subject.

In order to train the RJIVE, the AgeDB was divided into $M = 10$ age groups: '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 the FES task, 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 [43]. To this end, we compare the performance of RJIVE with the Illumination Aware Age Progression (IAAP) method [1], Coupled Dictionary Learning (CDL) method [2], Deep Ageing with Restricted Boltzmann

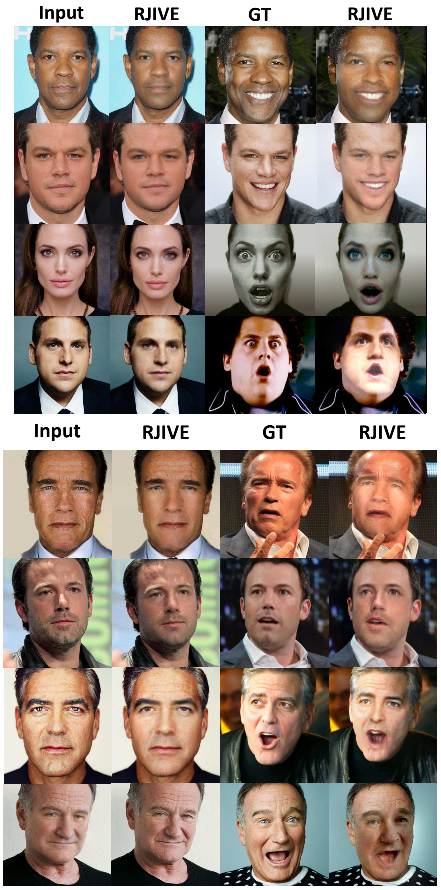


Fig. 6: Synthesized 'in-the-wild' expressions produced by the RJIVE method.

Machines (DARB) method [3], Craniofacial Growth (CG) [4] model, Exemplar-based Age Progression (EAP) [5] method, Face Transformer (FT Demo) [44], and Recurrent Face Aging (RFA) method [6]. In Figures 7, 8, 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. Figure 9[2] depicts faces synthesized by the DARB, IAAP, and RJIVE methods. By observing the results, it can be clearly seen that the identity information is not preserved in the case of DARB. In particular, the progressed faces of all subjects for a specific age group are very similar among them. It looks like all of them were created by transferring the skin colour from the input image to the same mean appearance. On the other hand, the identity information in the faces produced by the proposed RJIVE method remains. Finally, progressed examples faces in all of the age-groups produced by the RJIVE are visualized in Figure 10.

### 7.3.2 3D age progression of an unseen subject

Here the ability of the proposed RJIVE-M method to perform 3D face age progression is demonstrated. Similarly to the 2D face age progression experiments presented previously, the AgeDB database was divided into $M = 6$ age groups ('21-30', '31-40', '41-50', '51-60', '61-70', '71+') and used to train the RJIVE-M. In order to acquire the 3D training data for this task the 3DMM-ITW [30] was employed. The optimal shape and camera parameters were extracted by fitting the model to each one of the images of all age groups. In order to recover 3D shapes of high quality, we used the age and gender specific version of the LSFM shape model introduced in [45] in order to describe identity and the blendshapes of [46] in order to describe facial expressions. Having recovered the 3D shape of each face, we computed the self-occlusion mask by using ray-tracing (see first row of Figure 11). Then the completed joint and individual components of the grossly corrupted and incomplete UV textures were obtained by employing the RJIVE-M. The joint components obtained by applying a variant of JIVE with missing values, i.e., JIVE-M and the RJIVE-M on UV textures are displayed in Figure 11[2]. By observing the results, we can clearly see that the RJIVE successfully removed the occlusions produced by eyeglasses and fingers in all images. This is attributed to the fact that the matrix $\ell_1$-norm loss adopted in RJIVE, which effectively handles sparse noise of possibly large magnitude.

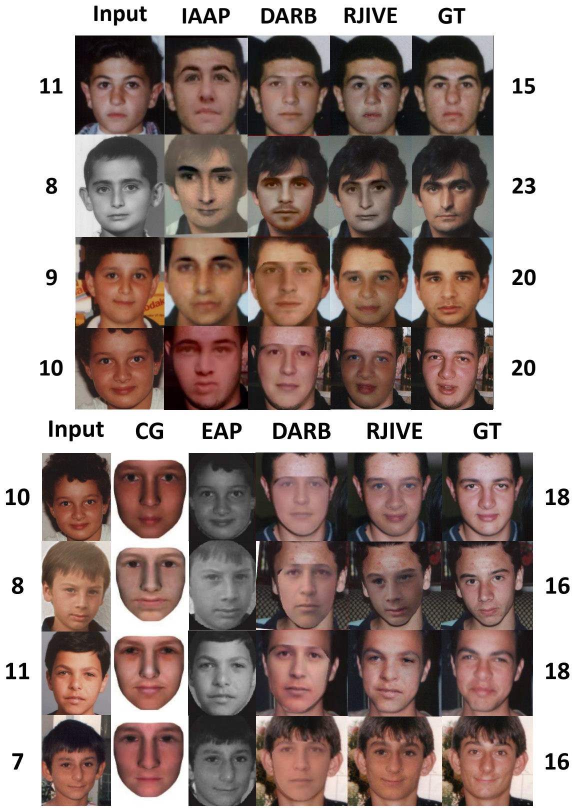


Fig. 7: Progressed faces produced by the compared methods on the FG-NET database.

Similarly, in the 2D face aging experiments we can apply the RJIVE-M to the recovered UV maps to learn components that can be used to age the UV texture of a test subject. Since, the 3D shapes are produced by the LSFM model they neither have missing values nor are contaminated by noise. Therefore, to find aging components for the 3D shape we used the standard JIVE.

In the test phase, the 3D shape of the test face is obtained by using the 3DMM-ITW algorithm [30]. Then the UV texture and the corresponding self-occlusion mask are computed by employing the recovered 3D shape. The progression of the texture of the test subject in an age group is obtained by solving the problem (13) (for the shape we use the problem in (10)). Progressed unseen subjects in all age groups, projected back in the image plane, are visualized in Figure 12[2]. Having calculated a progressed 3D texture image and a 3D shape, the resulting is projected back in

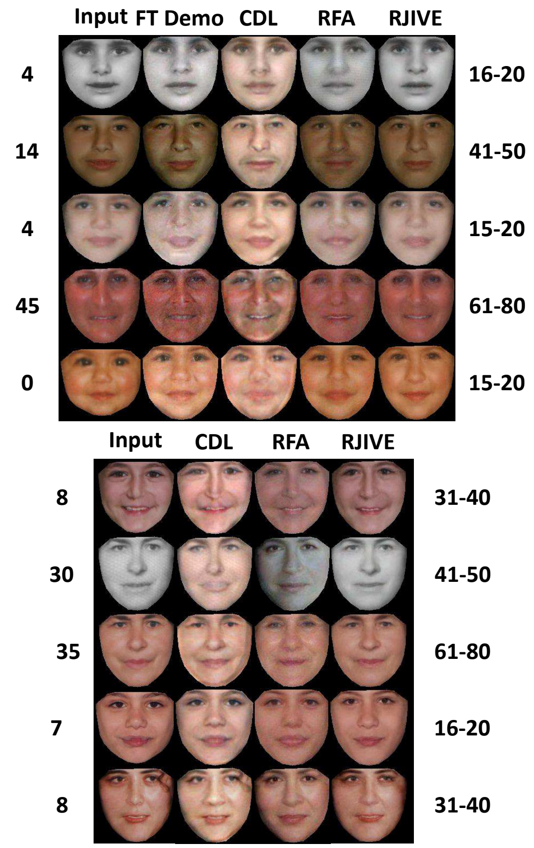


Fig. 8: Progressed faces produced by the compared methods on the FG-NET database.

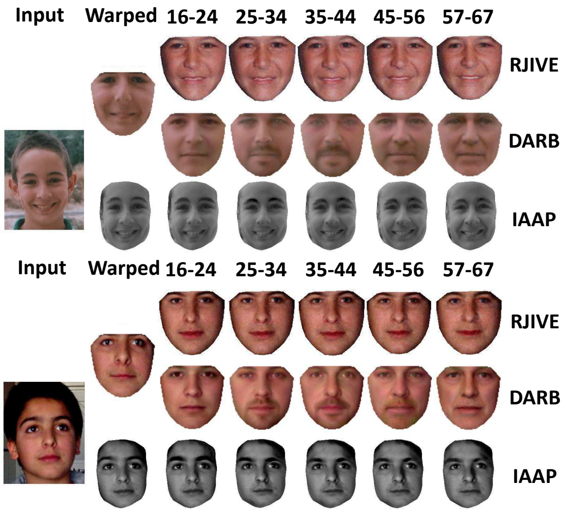


Fig. 9: Comparisons between the IAAP, DARB, and RJIVE methods.

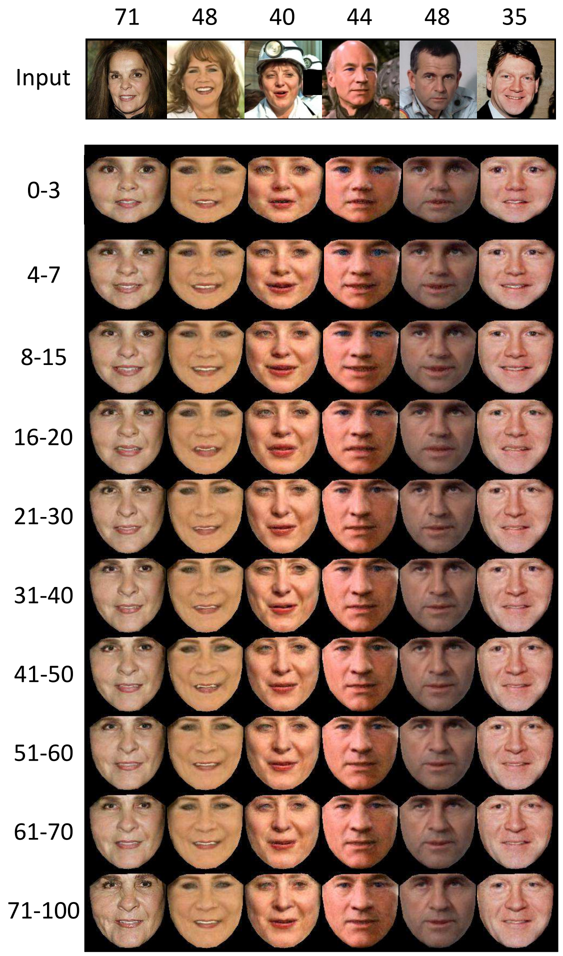


Fig. 10: Progressed faces produced by the proposed RJIVE method.

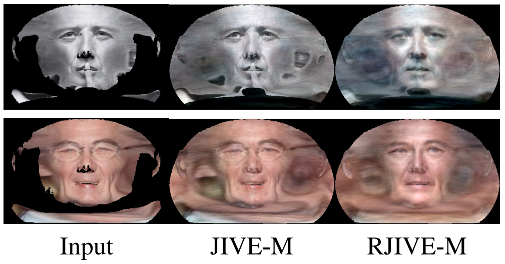


Fig. 11: Input images and corresponding joint components produced by the compared JIVE-M and RJIVE-M methods. As it can be observed the proposed method is able to remove occlusions produced by fingers and glasses.

the image plane using the camera parameters initially acquired by fitting the 3DMM-ITW in the test image.

Figure 13[2] presents additional results that demonstrate the ability of the RJIVE-M to perform not only age progression but also completion. For visualization purposes the completed and age-progressed 3D faces produced by the RJIVE-M were mapped on the progressed 3D shape. For each subject, the original and two side poses are depicted. The extracted by the 3DMM-ITW 3D face model of the input image is displayed in the first row. By observing the results it becomes apparent that due to self-occlusions, the instance of the 3D model with pose different to the input one, contains huge areas of missing values (black color). This is not the case for the progressed and completed results produced by the RJIVE-M (second row). As it can be seen, the completion of the regions with missing data is accurate and proves the significant representational power of the bases extracted by RJIVE-M.

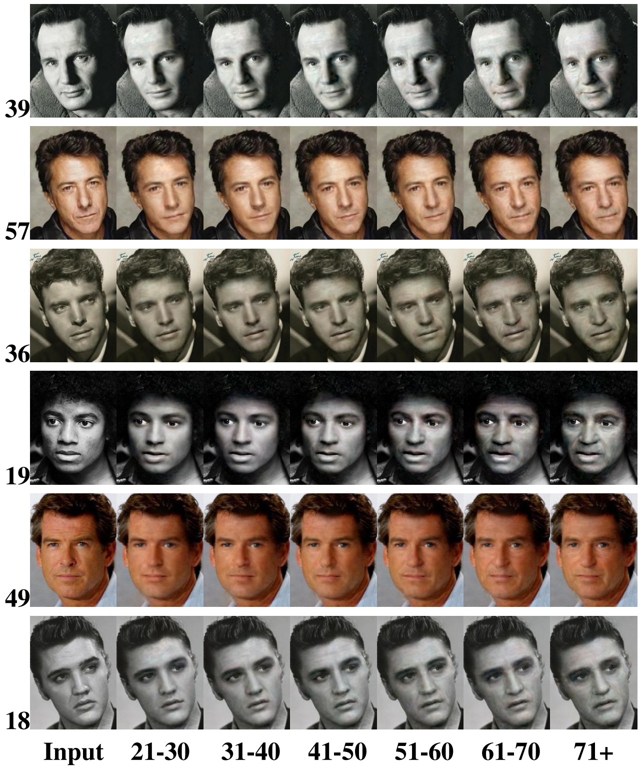


Fig. 12: Progressed faces produced by the proposed RJIVE-M, projected back in the image plane using mean 3D face shapes and camera parameters acquired by fitting the 3DMM-ITW.

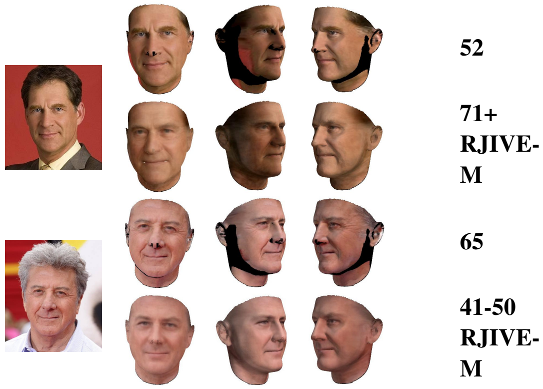
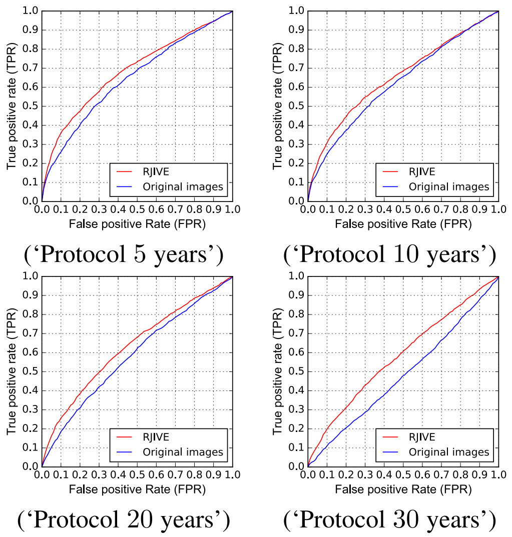


Fig. 13: Progressed and completed 3D texture images, produced by the proposed RJIVE-M method, mapped on mean 3D face shapes. The 3D face models are visualized in the original and two side poses, in order to make the differences between the completed and missing data visible.

### 7.3.3 Age-invariant face verification ‘in-the-wild’

The performance of the RJIVE is also quantitatively assessed by conducting age-invariant face verification experiments. Following the successfully used verification protocol of the LFW database [47], we propose four new age-invariant face verification protocols based on the 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 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., $\{5 \text{ years}, 10 \text{ years}, 20 \text{ years}, 30 \text{ years}\}$.



(‘Protocol 5 years’) (‘Protocol 10 years’)

(‘Protocol 20 years’) (‘Protocol 30 years’)

Fig. 14: ROC curves of RJIVE on the proposed protocols. ‘Original images’ corresponds to the results obtained by employing the actual images.

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 applied on their warped version in order to extract the joint and individual components. All images of each training pair were then progressed into $M = 10$ age groups resulting into 10 new pairs. The progressed images of six subjects are depicted in Figure 10. As we wanted to represent each pair by using a single feature, gradient orientations were extracted from the corresponding images and subsequently the mean value of their cosine difference was employed as the pair’s feature. $M$ different Support Vector Machines (SVM) were trained by utilizing the extracted features. Finally, the scores produced by all the SVMs were lately fused via an SVM.

In Figure 14, 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 Area Under Curve (AUC) are reported in Table 3. 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 large 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 5 difference is very large (e.g., 30 years).

TABLE 3: Mean AUC and Accuracy on the proposed protocols.

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

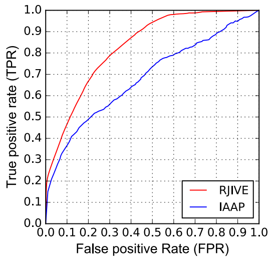


Fig. 15: ROC curve of the RJIVE and IAAP on FG-NET database.

The performance of RJIVE in age-invariant face verification is also compared against the IAAP [1] by conducting experiments on the FG-NET database. The experimental protocol employed is the following. By selecting images where the depicted subjects are older than the age of $18$ years, we created a subset of the FG-NET database consists of $518$ images. Then, based on the selected images we created $1250$ intra-class pairs, i.e., the images of each pair depict the same subject under different ages, and another $1250$ inter-class pairs. The experiment protocol was finally created by dividing the pairs on $5$ folds with each fold containing $250$ intra-class and $250$ inter-class pairs. All images were then progressed by employing the RJIVE and IAAP methods. A similar to the previous experiment procedure was followed in order to perform the age-invariant verification. The produced ROC curves are displayed in Figure 15. As it can be observed, the proposed RJIVE method outperforms the IAAP by a large margin, indicating that the RJIVE produces progressed images of high quality without removing the identity information.

## 8 Conclusions

A general framework for robust recovering of joint and individual variance among several datasets possibly contaminated by gross, non-Gaussian errors and missing values has been proposed in this paper. Four different models namely $\ell_1$-RJIVE, NN-$\ell_1$-RJIVE, SRJIVE, and RJIVE-M have been proposed. Furthermore, based on the recovered components from the training data, two novel optimization problems that extract the joint and individual components of an unseen test sample, are introduced. The effectiveness of the RJIVE was first tested by conducting experiments on synthetic data. Moreover, extensive experiments were conducted on facial expression synthesis and 2D and 3D face age progression by utilizing five datasets captured under both controlled and 'in-the-wild' conditions. The experimental results validate the effectiveness of the proposed RJIVE framework over the state-of-the-art.

## Acknowledgments

This work was partially funded by EPSRC Project EP/N007743/1 (FACER2VM). Finally, we would like to thank Alina Leidinger for some insights on an earlier version of this manuscript.

## References

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

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

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

[4] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proceedings of IEEE Intl Conference on Computer Vision & Pattern Recognition (CVPR)*, vol. 1, 2006, pp. 387–394.

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

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

[7] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[8] A. Klami and S. Kaski, "Probabilistic approach to detecting dependencies between data sets," *Neurocomputing*, vol. 72, no. 1, pp. 39–46, 2008.

[9] 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)*, vol. 36, no. 7, pp. 1299–1311, 2014.

[10] L. R. Tucker, "An inter-battery method of factor analysis," *Psychometrika*, vol. 23, no. 2, pp. 111–136, 1958.

[11] A. Klami, S. Virtanen, E. Leppäaho, and S. Kaski, "Group factor analysis," *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, vol. 26, no. 9, pp. 2136–2147, 2015.

[12] 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*, vol. 7, no. 1, p. 523, 2013.

[13] 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)*, vol. 11, pp. 2426–2439, 2015.

[14] Y. Panagakis, M. Nicolaou, S. Zafeiriou, and M. Pantic, "Robust correlated and individual component analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 38, no. 8, pp. 1665–1678, 2015.

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

[16] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust joint and individual variance explained," *Proceedings of IEEE Intl Conference on Computer Vision & Pattern Recognition (CVPR)*, 2017.

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

[18] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[19] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[20] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.

[21] 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*, vol. 59, no. 6, pp. 797–829, 2006.

[22] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[23] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[24] J. Wright and Y. Ma, "Dense error correction via $\ell_1$-minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3540–3560, 2010.

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

[26] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust statistical frontalization of human and animal faces," *Intl Journal of Computer Vision (IJCV)*, pp. 1–22, 2016.

[27] ——, "RAPS: Robust and efficient automatic construction of person-specific deformable models," in *Proceedings of IEEE Intl Conference on Computer Vision & Pattern Recognition (CVPR)*, 2014, pp. 1789–1796.

[28] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 11, pp. 2233–2246, 2012.

[29] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[30] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou, "3d face morphable models" in-the-wild"," *Proceedings of IEEE Intl Conference on Computer Vision & Pattern Recognition (CVPR)*, 2017.

[31] C. Sagonas, Y. Panagakis, S. Arunkumar, N. Ratha, and S. Zafeiriou, "Back to the future: A fully automatic method for robust age progression," in *Proceedings of IEEE Intl Conference on Pattern Recognition (ICPR)*, 2016, pp. 4226–4231.

[32] 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 Intl Conference on Multimedia, Open Source Software Competition*, Orlando, FL, USA, November 2014, pp. 679–682.

[33] 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)*, vol. 47, pp. 3–18, 2016.

[34] I. Matthews and S. Baker, "Active appearance models revisited," *Intl Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 135–164, 2004.

[35] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing (IMAVIS)*, vol. 28.

[36] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Proceedings of IEEE Intl Conference on Computer Vision & Pattern Recognition, Workshops (CVPR-W)*, 2010, pp. 94–101.

[37] 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)*. Springer, 2010, pp. 364–377.

[38] 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 Intl Conference on Computer Vision & Pattern Recognition, Workshops (CVPR-W)*, 2016.

[39] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, 2003, pp. 313–318.

[40] 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)*. Springer, 2014, pp. 768–783.

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

[42] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "Agedb: the first manually collected, in-the-wild age database," in *Proceedings of IEEE Intl Conference on Computer Vision & Pattern Recognition, Workshops (CVPR-W)*, 2017.

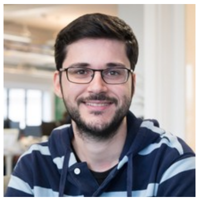
[43] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 24, no. 4, pp. 442–455, 2002.

[44] "Face transformer," http://morph.cs.standrews.ac.uk/Transformer/.

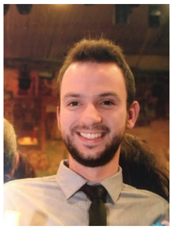
[45] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, "Large scale 3d morphable models," *Intl Journal of Computer Vision (IJCV)*, pp. 1–22, 2017.

[46] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "Facewarehouse: A 3d facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2014.

[47] 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*, pp. 14–003, 2014.



**Christos Sagonas** received the BSc and MSc degrees in Computer Science from Aristotle University of Thessaloniki, Greece in 2009 and 2011, respectively. In 2012 he joined the Intelligent Behaviour Understanding Group (IBUG) at Computing Department, Imperial College London, where he received the PhD degree in 2016. During his PhD, Christos worked mainly on machine-learning and computer vision models for automated facial landmark localization under totally unconstrained conditions. His current research interests include machine learning and computer vision with applications to human face analysis.



**Evangelos Ververas** graduated in September 2016 from the Department of Electrical and Computer Engineering in Aristotle University of Thessaloniki, in Greece. He joined the Intelligent Behavior Understanding Group (IBUG) at the Department of Computing, Imperial College London, in October 2016 and he is currently working as a PhD Student/Teaching Assistant under the supervision of Dr. Stefanos Zafeiriou. His research focuses on machine learning and computer vision models for 3D reconstruction and analysis of human faces.

**Yannis Panagakis** is a Lecturer (Assistant Professor equivalent) in Computer Science at Middlesex University London and a Research Fellow at the Department of Computing, Imperial College London. His research interests lie in machine learning and its interface with signal processing, high-dimensional statistics, and computational optimization. Specifically, Yannis is working on models and algorithms for robust and efficient learning from high-dimensional data and signals representing audio, visual, affective, and social information. He has been awarded the prestigious Marie-Curie Fellowship, among various scholarships and awards for his studies and research. Yannis currently serves as an Associate Editor of the Image and Vision Computing Journal. He co-organized the BMVC 2017 and several workshops and special sessions in top venues such as ICCV. He received his PhD and MSc degrees from the Department of Informatics, Aristotle University of Thessaloniki and his BSc degree in Informatics and Telecommunication from the University of Athens, Greece.

**Stefanos P. Zafeiriou** (M09) is currently a Reader in Machine Learning and Computer Vision with the Department of Computing, Imperial College London, London, U.K, and a Distinguishing Research Fellow with University of Oulu under Finish Distinguishing Professor Programme. He was a recipient of the Prestigious Junior Research Fellowships from Imperial College London in 2011 to start his own independent research group. He was the recipient of the President's Medal for Excellence in Research Supervision for 2016. He has received various awards during his doctoral and post-doctoral studies. He currently serves as an Associate Editor of the IEEE Transactions on Affective Computing and Computer Vision and Image Understanding journal. In the past he held editorship positions in IEEE Transactions on Cybernetics the Image and Vision Computing Journal. He has been a Guest Editor of over six journal special issues and co-organised over 13 workshops/special sessions on specialised computer vision topics in top venues, such as CVPR/FG/ICCV/ECCV (including three very successfully challenges run in ICCV13, ICCV15 and CVPR'17 on facial landmark localisation/tracking). He has co-authored over 55 journal papers mainly on novel statistical machine learning methodologies applied to computer vision problems, such as 2-D/3-D face analysis, deformable object fitting and tracking, shape from shading, and human behaviour analysis, published in the most prestigious journals in his field of research, such as the IEEE T-PAMI, the International Journal of Computer Vision, the IEEE T-IP, the IEEE T-NNLS, the IEEE T-VCG, and the IEEE T-IFS, and many papers in top conferences, such as CVPR, ICCV, ECCV, ICML. His students are frequent recipients of very prestigious and highly competitive fellowships, such as the Google Fellowship x2, the Intel Fellowship, and the Qualcomm Fellowship x3. He has more than 4500 citations to his work, h-index 36. He is the General Chair of BMVC 2017.