Multi-Attribute Robust Component Analysis for Facial UV Maps

Stylianos Moschoglou*, *Student Member, IEEE,* Evangelos Ververas*, Yannis Panagakis, *Member, IEEE,* Mihalis A. Nicolaou, *Member, IEEE,* and Stefanos Zafeiriou, *Member, IEEE*

***Abstract*—The collection of large scale 3D face models has led to significant progress in the field of 3D face alignment "in-the-wild", with several methods being proposed towards establishing sparse or dense 3D correspondences between a given 2D facial image and a 3D face model. Utilizing 3D face alignment improves 2D face alignment in many ways, such as alleviating issues with artefacts and warping effects in texture images. However, the utilization of 3D face models introduces a new set of challenges for researchers. Since facial images are commonly captured in arbitrary recording conditions, a considerable amount of missing information and gross outliers is observed (e.g. due to self-occlusion, subjects wearing eye-glasses and so on). To this end, in this paper we propose the Multi-Attribute Robust Component Analysis (MA-RCA), a novel technique that is suitable for facial UV maps containing a considerable amount of missing information and outliers, while additionally, elegantly incorporates knowledge from various available attributes, such as *age* and *identity*. We evaluate the proposed method on problems such as UV denoising, UV completion, facial expression synthesis and age progression, where MA-RCA outperforms compared techniques.**

***Index Terms*—Robust Component Analysis, Low Rank, Sparsity, Facial UV Maps.**

I. Introduction

Significant progress has been observed during the past years in the field of sparse and dense 3D face alignment [2], [10], [12], [24], [25]. Recent developments include the utilization of Deep Neural Networks (DNNs) for estimation of 3D facial structure, as well as a methodology for fitting a 3D Morphable Model (3DMM) in "in-the-wild" images [2]. Additionally, several benchmarks for training sparse 3D face alignment models have been recently developed [2], [24]. The utilization of these methods introduces new challenges and opportunities as far as facial texture is concerned.

In particular, by sampling over the fitted image, a 2D UV map of the facial texture can be constructed. A UV texture map is the main way to store the texture of a 3D model as a simple 2D image in graphics applications. In more detail, it is a 2D representation of the texture of a 3D model which is precisely connected with the object's vertices through a set of coordinates that map each vertex to a specific location of the UV map. The mapping coordinates naturally occur by unwrapping the 3D model and scaling it to the UV texture map's space [4]. In our work we employ UV texture maps

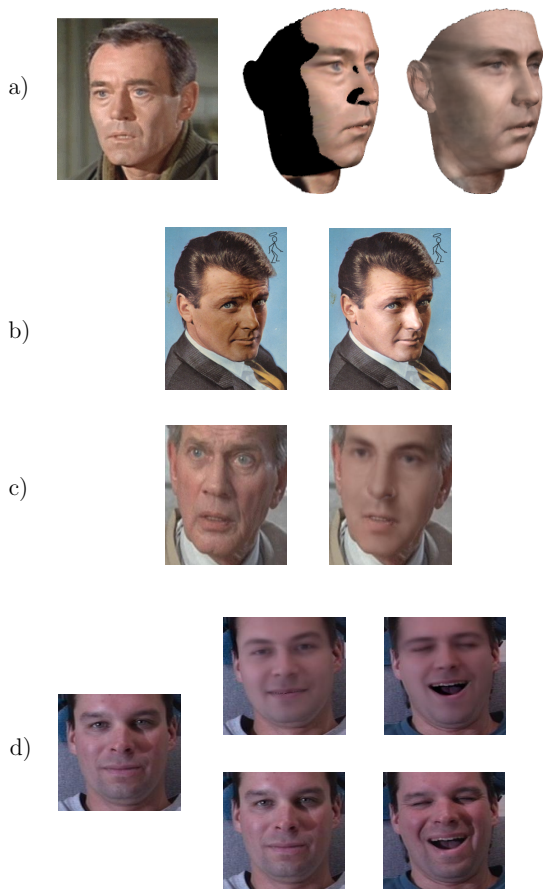

Fig. 1. Various indicative applications of MA-RCA, such as: a) facial UV map completion: On the left is the original image, in the middle is the extracted 3D facial shape with the original texture which has missing values, on the right is the 3D facial shape with the reconstructed completed facial texture b) illumination transfer: on the left is the original image, on the right is the illuminated one, c) age progression: on the left is the original image (old), on the right is the age-progressed one (young), d) facial expression synthesis: on the left is the neutral image, in the top row are the reconstructed facial expressions (smile and scream) from the neutral one and in the bottom row are the actual, ground truth facial expressions (smile and scream).

*S. Moschoglou and E. Ververas contributed equally to this work.
S. Moschoglou, E. Ververas, Y. Panagakis, and S. Zafeiriou are with the Department of Computing, Imperial College London, United Kingdom.
E-mail: {s.moschoglou, e.ververas16, i.panagakis, s.zafeiriou}@imperial.ac.uk
M. A. Nicolaou is with the Computation-based Science and Technology Research Centre, The Cyprus Institute, and the Dept of Computing, Goldsmiths, University of London.
E-mail: m.nicolaou@cyi.ac.cy
Y. Panagakis is also with the Department of Computing, Middlesex University, London, UK
E-mail: y.panagakis@mdx.ac.uk
This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. The material includes extra quantitative as well as qualitative experiments. Contact {s.moschoglou, e.ververas16}@imperial.ac.uk for further questions about this work.

a) b) c)

Fig. 2. Original facial image is depicted in a). The facial image produced by a 2D face alignment algorithm is presented in b). The UV map derived by fitting a 3DMM is depicted in c).

that are precisely mapped to 3D models of human faces and particularly to instances of the LSFM face model [3]. An example of a UV map with missing values along with the original face image and the recovered shape projected on it, is depicted in Fig. 4.

A facial UV map may contain a considerable amount of missing data (pixels) due to factors such as self-occlusion. Nevertheless, it does not suffer from warping effects, in contrast to the facial images produced by a 2D face alignment algorithm, as can be clearly seen in Fig. 2. Utilizing facial UV maps for the discovery of latent components suitable for specific tasks (such as age or illumination transfer) requires the design of statistical component analysis methods that (a) can appropriately handle missing values, (b) can alleviate problems arising from gross errors, and (c) exploit any existing labels/attributes that are available. To tackle the aforementioned issues, it is natural to adopt techniques from the family of robust component analysis.

In the past years, significant research has been conducted in terms of formulating robust component analysis techniques. Arguably, the most prominent example lies in the Robust PCA (RPCA) algorithm [5], that has also been extended for handling missing values in [21]. The RPCA algorithm with missing values has been recently proven extremely useful towards the extraction of a low-rank sub-space of facial UV textures that is free of gross errors, thus deeming it extremely useful for the fitting of 3DMMs "in-the-wild" [2]. Nevertheless, RPCA is an unsupervised component analysis technique, and hence does not take into account the various attributes/annotations that may be present in the data at-hand.

Different variations of RPCA [5] have been introduced in the literature, such as the Robust Principal Component Analysis with Non-Greedy $l1$-Norm Maximization (RPCA-L1) [15] and the Optimal Mean Robust Principal Component Analysis (OM-RPCA) [16], both of which propose different methods for solving the RPCA optimization problem ([5]). In RPCA-L1 [15], the authors first propose an efficient optimization algorithm to solve a general $l1$-norm maximization problem and then employ it to formulate a non-greedy solution of the $l1$-norm RPCA problem. OM-RPCA [16] solves the traditional $l1$-norm RPCA optimization problem while simultaneously updating the mean component of the data matrix at each iteration of the algorithm, aiming to achieve better separation of the data. Similarly to the common RPCA algorithm [5], none of the two aforementioned methods are designed to exploit any available labels or annotations of the data.

Other recent robust component analysis methods include Robust Correlated and Individual Component Analysis (RCICA) [17], as well as Robust Joint and Individual Variance Explained (RJIVE) [19]. RCICA robustly recovers both the correlated and individual low-rank components of two views of noisy data, and can therefore be interpreted as a robust extension of Canonical Correlation Analysis (CCA) [23]. Nevertheless, RCICA is not designed to utilize labels or any available annotations. RJIVE further extends RCICA by extracting low-rank sub-spaces from multiple-views similarly to RCICA in the presence of a *single* attribute only (e.g., if the data at-hand are annotated for the attribute *age*, then they may be split in different age-groups and each age-group can be considered as a different view). As a result, data that is annotated in terms of multiple attributes (such as *identity* and *age*) cannot be fully exploited in RJIVE.

To alleviate the shortcomings of the previously mentioned methods, in this paper we introduce Multi-Attribute Robust Component Analysis, dubbed MA-RCA, which takes a much more natural approach to the problem of facial UV analysis, since it is able to inherently incorporate the existence of multiple attributes at-hand during the training as well as the testing processes. An example of how MA-RCA decomposes a UV map into the various components is presented in Fig. 3. In summary, the contributions of the paper are as follows.

- We introduce MA-RCA, a novel component analysis technique which recovers suitable components that robustly capture the shared and individual variation of data under a multi-attribute scenario. Furthermore, MA-RCA is inherently able to handle observations with missing values, as well as sparse and gross corruptions.
- We demonstrate that MA-RCA can be applied to a number of challenging problems, such as completion of missing data in the texture of a reconstructed 3D facial image and *transfer* of multiple attributes in images captured "in-the-wild" (e.g., *illumination*, *identity*, *expression* and *age*).

The rest of the paper is organized as follows. In Section II we provide the mathematical formulation of MA-RCA and present all the necessary optimization algorithms. In Section III we run a series of experiments and demonstrate the merits of MA-RCA against other state-of-the-art algorithms.

II. Multi-Attribute Robust Component Analysis

A. Preliminaries

Prior to delving into the model, a few explanations regarding the notations used throughout the paper are provided. Lower-case letters, e.g., x, denote scalars, lower-case (upper-case) bold letters denote vectors (matrices), e.g., $\mathbf{x}$ ($\mathbf{X}$). Moreover, L_1 (L_2) vector norm is defined as $\|\mathbf{x}\|_1 \doteq \sum_i |x_i|$ ($\|\mathbf{x}\|_2 \doteq \sqrt{\sum_i x_i^2}$). Similarly, $L_{1,1} \equiv L_1$ ($L_{2,2} \equiv L_F$) matrix norm is defined as $\|\mathbf{X}\|_1 \doteq \sum_{i,j} |x_{ij}|$ ($\|\mathbf{X}\|_F \doteq \sqrt{\sum_{i,j} x_{ij}^2}$). The nuclear norm of a matrix $\mathbf{X}$, i.e., the sum of its singular values, is defined as $\|\mathbf{X}\|_*$. The Hadamard, i.e., element-wise, product of two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted as $\mathbf{X} \odot \mathbf{Y}$. Finally,

Fig. 3. Facial UV map decomposition utilizing MA-RCA. Vectorized UV map $\mathbf{x}_{30}$ is decomposed into: a) an *age-group* component $\mathbf{F}_1\mathbf{h}_{1,3}$, where $\mathbf{F}_1$ is the base that renders the age-groups and $\mathbf{h}_{1,3}$ is the shared age selector corresponding to the third age-group in the training set (in this example, the age-group is *41-50*), b) an *identity* component $\mathbf{F}_2\mathbf{h}_{2,5}$, where $\mathbf{F}_2$ is the base that renders the identities and $\mathbf{h}_{2,5}$ is the shared identity selector corresponding to the fifth identity in the training set (in this example, the identity is *Frank Sinatra*), c) an individual component $\mathbf{G}_{30}$ that captures all of the error-free information of the particular UV map that cannot be explained by the *age* and *identity* components, d) an error component $\mathbf{E}_{30}$ that captures the gross errors of the particular UV map (in this example, occlusions which correspond to sparse, non-Gaussian noise).

(a) (b) (c) (d)

Fig. 4. Extraction of UV texture map with missing values from a 2D image using a face 3DMM. Having recovered the (b) 3D shape of a face by fitting a 3DMM into (a) an input image we can (c) project the reconstructed 3D shape on it and then compute the UV texture map along with the occluded parts.

we provide the following operator definitions which will be utilized in the mathematical derivations required for MA-RCA.

- Procrustes operator: $\mathcal{Q}(\mathbf{X}) \doteq \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are given by the rank-r Singular Value Decomposition (SVD) of $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.
- Shrinkage operator: $\mathcal{S}_\tau(\sigma) \doteq \text{sgn}(\sigma)\max(|\sigma| - \tau, 0)$
- Singular Value Thresholding (SVT) operator: $\mathcal{D}_\tau(\mathbf{X}) \doteq \mathbf{U}\text{diag}(\mathcal{S}_\tau(\mathbf{d}))\mathbf{V}^T$, where $\mathbf{X} = \mathbf{U}\text{diag}(\mathbf{d})\mathbf{V}^T$ is the SVD of $\mathbf{X}$.

B. Problem formulation

Without any loss of generality, suppose that the incomplete, contaminated by gross errors UV maps at-hand are annotated for J attributes (e.g., *identity*, *age*, etc.), where each attribute may have $M_i, \forall i \in \{1,\dots,J\}$, different instantiations (e.g., attribute *identity* may have the instantiation *Frank Sinatra*, *Albert Einstein*, etc.). Moreover, assume that there is a total of N samples in the training set. Aim of MA-RCA is to robustly extract J joint components corresponding to the available attributes during training, an individual component which captures the rest data information that cannot be explained by the J components and a component which captures the gross but sparse errors. Let training data be concatenated in a column-wise manner, i.e., $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_N]$, where $\mathbf{x}_i \in \mathbb{R}^{F\times 1}, i \in \{1,\dots,N\}$, is a vectorized form of a facial UV map. Then MA-RCA admits the following decomposition.

$$\mathbf{X} = \sum_{i=1}^{J} \mathbf{S}_i + \mathbf{G} + \mathbf{E}, \tag{1}$$

where $\mathbf{S}_i, i \in \{1,\dots,J\}$, are the J shared components for every attribute, $\mathbf{G}$ is the individual component and $\mathbf{E}$ is the error component.

Nevertheless, $\mathbf{S}_i, i \in \{1,\dots,J\}$, must have a specific low-rank structure which accounts for the different instantiations of each attribute. That is, every attribute should be rendered by a *base* and subsequently every corresponding instantiation be rendered by a *selector* on that base. Therefore, (1) is re-formulated as follows.

$$\mathbf{X} = \sum_{i=1}^{J} \mathbf{F}_i\mathbf{H}_i + \mathbf{G} + \mathbf{E}, \tag{2}$$

where $\mathbf{F}_i \in \mathbb{R}^{F\times M_i}, i \in \{1,\dots,J\}$, are the *bases* that render each attribute and $\mathbf{H}_i \in \mathbb{R}^{M_i\times N}, i \in \{1,\dots,J\}$, are comprised of the *shared selectors*, i.e., $\mathbf{H}_i \doteq [\mathbf{h}_{i,a} \dots \mathbf{h}_{i,b}]$ (such that $a, b \in \{1,\dots,M_i\}$), which render a specific instantiation for an attribute (e.g., assuming that base $\mathbf{F}_i$ renders attribute *identity*, then $\mathbf{h}_{M_i}$ would render a particular instantiation of this attribute, e.g., *Albert Einstein*). It should be noted that data which bear the same instantiation (e.g., M_i) for a particular attribute (e.g., attribute i) have the same selector $\mathbf{h}_{i,M_i}$ (for instance, multiple data with instantiation *Albert Einstein* will all have the selector $\mathbf{h}_{i,M_i}$). Furthermore, low-rank base $\mathbf{G} \in \mathbb{R}^{F\times N}$, renders the individual variation for all of the images in the training set that cannot be explained by the existing attributes. Finally, $\mathbf{E} \in \mathbb{R}^{F\times N}$ encapsulates gross errors (such as occlusions, pixel corruptions, etc.) for all of the data samples in the training set.

In order to recover components $\{\mathbf{F}_i\mathbf{H}_i\}_{i=1}^{J}$ and $\mathbf{G}$ which are as informative as possible, the error term which accounts for the existence of gross but sparse errors in the visible parts of the UVs has to be minimized. This is equivalent to minimizing the L_0 norm of the error term for the visible parts of the UV maps. However, to avoid the NP-hardness of the L_0 norm minimization we adopt the L_1 norm as the tightest convex surrogate [7].

The problem is then formulated as follows.

$$\begin{aligned} &\min_{\theta} \|\mathbf{W} \odot \mathbf{E}\|_1, \\ &\text{s.t. } \mathbf{X} = \sum_{i=1}^{J} \mathbf{F}_i \mathbf{H}_i + \mathbf{G} + \mathbf{E} \\ &\quad \left\{\mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}\right\}_{i=1}^{J}, \text{rank}(\mathbf{G}) = R, \end{aligned} \tag{3}$$

where $\theta = \{\mathbf{F}_i, \mathbf{H}_i, \mathbf{G}\}, i \in \{1, \ldots, J\}$ and $R, R < \min(F, N)$, is a hyper-parameter. Moreover, $\mathbf{W} \doteq \left[\mathbf{w}_1 \ldots \mathbf{w}_N\right]$, where $\mathbf{w}_i \in \{0,1\}^{F \times 1}, i\{1, \ldots, N\}$, is the corresponding vectorized occlusion mask for each UV map $\mathbf{x}_i$. The visible (missing) pixels for each UV map correspond to ones (zeros) in each matching occlusion mask. Orthonormalization constraints on the bases $\{\mathbf{F}_i\}_{i=1}^{J}$ facilitate the recovery of unique and identifiable selectors. Because of the fact that R is a hyper-parameter, it requires a large number of experiments to estimate the optimal rank for $\mathbf{G}$. Since R is upper bounded, the following relaxed decomposition can be used to automatically recover the optimal rank for $\mathbf{G}$.

$$\begin{aligned} &\min_{\theta} \lambda \|\mathbf{W} \odot \mathbf{E}\|_1 + \|\mathbf{G}\|_*, \\ &\text{s.t. } \mathbf{X} = \sum_{i=1}^{J} \mathbf{F}_i \mathbf{H}_i + \mathbf{G} + \mathbf{E}, \\ &\quad \left\{\mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}\right\}_{i=1}^{J}, \end{aligned} \tag{4}$$

where the nuclear norm of $\mathbf{G}$ is introduced as a convex surrogate of the rank function [5] and $\lambda > 0$ is a regularizer.

Replacing the parameter R with the regularizer λ makes the problem much easier to address and in much less time [5], [19]. Regularizer λ can be selected by using the formula $\lambda = 1/\sqrt{max(F, N)}$ (where F, N correspond to the rows and columns of the data matrix, respectively), as shown in [5], [19]. In the supplemental material we further corroborate this by conducting a number of different experiments on various values of λ.

C. Mathematical derivations

Because problem (4) is separable, we adopt an alternating optimization scheme to find the updates for every parameter. The corresponding partially Augmented Lagrangian for (4) may then be written as

$$\begin{aligned} \mathcal{L}(\theta) = &\lambda \|\mathbf{W} \odot \mathbf{E}\|_1 + \|\mathbf{G}\|_* - \frac{1}{2\mu}\|\mathbf{\Lambda}\|_F^2 + \\ &+ \frac{\mu}{2}\left\|\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i \mathbf{H}_i - \mathbf{G} - \mathbf{E} + \frac{\mathbf{\Lambda}}{\mu}\right\|_F^2, \\ &\text{s.t. } \left\{\mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}\right\}_{i=1}^{J}, \end{aligned} \tag{5}$$

where $\theta = \{\mathbf{F}_i, \mathbf{H}_i, \mathbf{G}, \mathbf{\Lambda}\}, i \in \{1, \ldots, J\}$. Problem (5) is minimized by employing the Alternating Direction Method of Multipliers (ADMM) [1], [8]. The algorithm for solving (5) is presented in Algorithm 1. The algorithm terminates when the iterations reach a predefined maximum value or a convergence criterion is met. The convergence criterion is met when the normalized reconstruction error, i.e., $\left\|\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i[t]\mathbf{H}_i[t] - \mathbf{G}[t] - \mathbf{W} \odot \mathbf{E}[t]\right\|_F / \|\mathbf{X}\|_F$ is less than a predefined threshold ϵ. The ADMM iteration reads as follows.

Update the primal variables:

For obtaining $\mathbf{H}_i, i \in \{1, \ldots, J\}$, where, as previously mentioned, $\mathbf{H}_i = \left[\mathbf{h}_{i,a} \ldots \mathbf{h}_{i,b}\right]$, we need to solve individually for every $\mathbf{h}_{i,j}, j \in \{1, \ldots, M_i\}$. Based on (5), the solution is given by minimizing

$$\begin{aligned} \mathbf{h}_{i,j}[t+1] = \underset{\mathbf{h}_{i,j}[t]}{\text{argmin}} \Bigg\| &\mathbf{X} - \sum_{k=1}^{J} \mathbf{F}_k[t]\mathbf{H}_k[t] - \\ &- \mathbf{G}[t] - \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg\|_F^2. \end{aligned} \tag{6}$$

Problem (6) admits a closed-form solution, which is

$$\begin{aligned} \mathbf{h}_{i,j}[t+1] = \frac{1}{N_{i,j}}\left(\mathbf{F}_i^{i,j}[t]\right)^T \Bigg(&\mathbf{X}^{i,j} - \sum_{k=1, k\neq i}^{J} \mathbf{F}_k^{i,j}[t]\mathbf{H}_k^{i,j}[t] - \\ &\mathbf{G}^{i,j}[t] - \mathbf{E}^{i,j}[t] + \frac{\mathbf{\Lambda}^{i,j}[t]}{\mu[t]}\Bigg) \cdot \mathbf{1}, \end{aligned} \tag{7}$$

where superscript $\{i, j\}$ means that only the $N_{i,j}$ columns corresponding each time to the j-th instantiation of the i-th attribute are considered (e.g., columns corresponding to data annotated for attribute *identity* and instantiation *Albert Einstein*) and $\mathbf{1}$ is a column vector of $N_{i,j}$ ones.

For deriving subspace $\mathbf{F}_i, i \in \{1, \ldots, J\}$, the following needs to be solved.

$$\begin{aligned} \mathbf{F}_i[t+1] = &\underset{\mathbf{F}[t]}{\text{argmin}} \left\|\mathbf{X} - \sum_{j=1}^{J} \mathbf{F}_j[t]\mathbf{H}_j[t+1] - \mathbf{G}[t] - \right. \\ &\left. - \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\right\|_F^2, \\ &\text{s.t. } \left\{\mathbf{F}_i^T \mathbf{F}_i = \mathbf{I}\right\}_{i=1}^{J}. \end{aligned} \tag{8}$$

In order to solve (8), we rely on the Procrustes Operator $\mathcal{Q}$ and the Lemma introduced next.

Lemma: The constraint minimization problem

$$\begin{aligned} \mathbf{\Omega}^* = &\underset{\mathbf{\Omega}}{\text{argmin}} \|\mathbf{\Omega}\mathbf{A} - \mathbf{B}\|_F^2 \\ &\text{s.t. } \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I}, \end{aligned} \tag{9}$$

has a closed-form solution [20] of the form $\mathbf{\Omega}^* = \mathcal{Q}\left(\mathbf{B}\mathbf{A}^T\right)$. As a result, the solution for (8), taking into account (9), is

$$\begin{aligned} \mathbf{F}_i[t+1] = \mathcal{Q}\Bigg[\Bigg(&\mathbf{X} - \sum_{j=1, j\neq i}^{J} \mathbf{F}_j[t]\mathbf{H}_j[t+1] - \mathbf{G}[t] - \\ &- \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg) \cdot \mathbf{H}_i[t+1]^T\Bigg]. \end{aligned} \tag{10}$$

For obtaining subspace $\mathbf{G}$, the following needs to be solved.

$$\begin{aligned} \mathbf{G}[t+1] = \underset{\mathbf{G}[t]}{\text{argmin}} \Bigg[\frac{\mu[t]}{2}\Bigg\|\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \\ - \mathbf{G}[t] - \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg\|_F^2 + \|\mathbf{G}[t]\|_*\Bigg]. \end{aligned} \tag{11}$$

Problem (11) is solved utilizing the SVT operator $\mathcal{D}$. The solution is

$$\mathbf{G}[t+1] = \mathcal{D}_{\frac{1}{\mu[t]}}\Bigg[\mathbf{X} - \sum_{i=1}^{J}\mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \mathbf{E}[t] + \\ + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg]. \tag{12}$$

For obtaining $\mathbf{E}$, the following problem needs to solved.

$$\mathbf{E}[t+1] = \underset{\mathbf{E}[t]}{\operatorname{argmin}}\Bigg[\lambda\left\|\mathbf{W}\odot\mathbf{E}[t]\right\|_1 + \frac{\mu[t]}{2}\Bigg\|\mathbf{X} - \\ - \sum_{i=1}^{J}\mathbf{F}_i[t+1]\cdot\mathbf{H}_i[t+1] - \mathbf{G}[t+1] - \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg\|_F^2\Bigg] \tag{13}$$

Problem (13) is solved utilizing the Shrinkage operator $\mathcal{S}$. The solution is

$$\mathbf{E}[t+1] = \mathbf{W}\odot\mathbf{Y} + \overline{\mathbf{W}}\odot\Bigg[\mathbf{X} - \sum_{i=1}^{J}\mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \\ - \mathbf{G}[t+1] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg], \tag{14}$$

where

$$\mathbf{Y} = \mathcal{S}_{\frac{\lambda}{\mu[t]}}\Bigg[\mathbf{X} - \sum_{i=1}^{J}\mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \\ - \mathbf{G}[t+1] + \frac{\mathbf{\Lambda}[t]}{\mu[t]}\Bigg] \tag{15}$$

and $\overline{\mathbf{W}}$ is the complement of $\mathbf{W}$.

Update the Lagrange multiplier and μ_t parameter:

$$\mathbf{\Lambda}[t+1] = \mathbf{\Lambda}[t] + \mu[t]\Bigg(\mathbf{X} - \sum_{i=1}^{J}\mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \\ - \mathbf{G}[t+1] - \mathbf{E}[t+1]\Bigg), \tag{16}$$

$$\mu[t+1] = \min\left(\rho\mu[t], \mu_{\max}\right). \tag{17}$$

Regarding the theoretical convergence of the ADMM algorithm presented previously, there is no proof when ADMM is utilized in settings with more than two blocks of variables. Nevertheless, ADMM provides good results in non-linear optimization problems [18]. Experimental evaluation of MA-RCA in a number of different tasks on various facial datasets admits that the derived solutions constitute a good approximation. It should be noted that despite using convex surrogates for the L_0 norm and the rank function, the problem is still non-convex due to the product terms of the MA-RCA formulation. Non-convex relaxation techniques could be utilized in a future work.

The time complexity for MA-RCA at each iteration of Algorithm 1 is $\sum_{i=1}^{J} O\left(\max\left(M_i F^2\right)\right) + O\left(\left(\max\left(NF^2, F^2N\right)\right) = O\left(\max\left(FN^2, F^2N\right)\right)\right.$, due to SVD applied on both the shared J components and the individual one and since $M_i < \min(F,N), \forall i \in \{1,\ldots,J\}$. The memory complexity of MA-RCA is $O\left(FN\right)$, due to SVD.

As mentioned previously, the type of noise which is mostly prevalent in facial UV maps is sparse, gross, non-Gaussian noise, due to occlusions, subjects wearing eye-glasses, pixel corruptions, etc. This is why MA-RCA is formulated to explicitly account for the sparse noise, as shown in (4). Furthermore, by increasing the regularizer λ (i.e., by reducing the effect of the sparse error term in (4)), MA-RCA will be able to handle data contaminated by both Gaussian and sparse noise. Finally, MA-RCA is able to handle data solely contaminated by Gaussian noise by vanishing the sparse error term, i.e., by setting $\lambda \to \infty$. This is further corroborated in Section III-A and Section III-B.

D. Reconstruction of a test image

After the optimal bases and the selectors have been recovered as described in Section II-C, they may be utilized in order to recover the shared and individual components of a test image. Then, the said components can be utilized in experiments such as completion of missing UV parts and *joint* transfer of a facial test image to another *age*, *identity*, *expression* or *illumination*, as demonstrated in Section III.

Without any loss of generality, assume a test UV map $\mathbf{y}$, which may be decomposed in the shared and individual components as follows.

$$\mathbf{y} = \sum_{i=1}^{J}\mathbf{F}_i\hat{\mathbf{h}}_i + \mathbf{K}\hat{\mathbf{w}} + \hat{\epsilon}, \tag{18}$$

where $\mathbf{K}$ is the linear span of $\mathbf{G}$, given by applying the rank-r SVD on $\mathbf{G}$. In the most general case, optimal selectors $\left\{\hat{\mathbf{h}}_i\right\} =_{i=1}^{J}$ and $\hat{\mathbf{w}}$ must be extracted by minimizing the sparse error term $\hat{\boldsymbol{\epsilon}}$ corresponding to the visible part of $\mathbf{y}$, for already recovered $\{\mathbf{F}_i\}_{i=1}^{J}$ and $\mathbf{G}$. That is, the following needs to be solved.

$$\min_{\theta}\left\|\mathbf{w}_y\odot\hat{\boldsymbol{\epsilon}}\right\|_1 \\ \text{s.t.}\quad \mathbf{y} = \sum_{i=1}^{J}\mathbf{F}_i\hat{\mathbf{h}}_i + \mathbf{K}\hat{\mathbf{w}} + \hat{\boldsymbol{\epsilon}}, \\ \hat{\mathbf{y}} = \sum_{i=1}^{J}\mathbf{F}_i\hat{\mathbf{h}}_i + \mathbf{K}\hat{\mathbf{w}}, \tag{19}$$

where $\mathbf{w}_y$ is the occlusion mask corresponding to the test UV map $\mathbf{y}$ and $\hat{\mathbf{y}}$ is the reconstructed, error-free facial UV map. In the case where, e.g., transfer of a test image to a specific age is required, the selector corresponding to the specific age will be fixed (i.e., the corresponding optimal selector found during the training process in Section II-C is utilized). Because problem (19) is separable, we adopt an alternating optimization scheme to find the updates for every

Algorithm 1: ADMM solver for problem (5)

Input: $\mathbf{X}$, where $\mathbf{X}$ is the set of the training data, $M_i, i \in \{1,\dots,J\}$, where M_i is number of instantiations for every attribute i, $\mathbf{W}$, where $\mathbf{W}$ is the set of occlusion masks corresponding to the training data and $\mathbf{L}$, where $\mathbf{L}$ is a one hot representation matrix of the instantiations assigned to each image over all the available attributes.

Initializations: $t=0, \epsilon = 1e-8$, maximum number of iterations $t_{\max}, \mu[0] = \frac{25}{\|\mathbf{X}\|}$, $\lambda = 1/\sqrt{max(F,N)}$ $\{\mathbf{F}_i[0], \mathbf{H}_i[0], \mathbf{G}[0], \mathbf{\Lambda}[0]\} = \mathbf{0}, i \in \{1,\dots,J\}, \rho = 1.2, \mu_{\max} = 10^7, \{\mathbf{F}_i[0]\}_{i=1}^{J} =$ random initializations so that $\mathbf{F}_i[0], i \in \{1,\dots,J\}$, are orthonormal.

Output: $\{\mathbf{F}_i[T], \mathbf{H}_i[T]\}_{i=1}^{J}, \mathbf{G}[T], \mathbf{E}[T]$.

while *not converged* **do**

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

    Update $\mathbf{H}_i$:

    for $j = 1 : M_i$ **do**

$$\mathbf{h}_{i,j}[t+1] = \tfrac{1}{N_{i,j}} \left(\mathbf{F}_i^{i,j}[t]\right)^T \left(\mathbf{X}^{i,j} - \sum_{k=1,k\neq i}^{J} \mathbf{F}_k[t]\mathbf{H}_k^{i,j}[t] - \mathbf{G}^{i,j}[t] - \mathbf{E}^{i,j}[t] + \frac{\mathbf{\Lambda}^{i,j}[t]}{\mu[t]} \right) \cdot \mathbf{1}$$

    end

$$\mathbf{H}_i[t+1] = \Big[\underbrace{\mathbf{h}_{i,1}[t+1] \dots \mathbf{h}_{i,a}[t+1]}_{N_{i,a} \text{ times}} \dots \underbrace{\mathbf{h}_{i,b}[t+1] \dots \mathbf{h}_{i,b}[t+1]}_{N_{i,b} \text{ times}} \Big] \text{ s.t. } a, b \in \{1,\dots,M_i\}$$

  end

  Update $\mathbf{F}$:

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

$$\mathbf{F}_i[t+1] = \mathcal{Q} \left[\left(\mathbf{X} - \sum_{j=1,j\neq i}^{J} \mathbf{F}_i[t]\mathbf{H}_i[t+1] - \mathbf{G}[t] - \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]} \right) \mathbf{H}_i^T[t+1] \right]$$

  end

  Update $\mathbf{G}$:

$$\mathbf{G}[t+1] = \mathcal{D}_{1/\mu[t]} \left[\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \mathbf{E}[t] + \frac{\mathbf{\Lambda}[t]}{\mu[t]} \right]$$

  Update $\mathbf{E}$:

$$\mathbf{E}[t+1] = \mathcal{S}_{\frac{\lambda}{\mu[t]}} \left[\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \mathbf{G}[t+1] + \frac{\mathbf{\Lambda}[t]}{\mu[t]} \right]$$

$$\mathbf{E}[t+1] = \mathbf{W} \odot \mathbf{E}[t+1] + \overline{\mathbf{W}} \odot \left[\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \mathbf{G}[t+1] + \frac{\mathbf{\Lambda}[t]}{\mu[t]} \right]$$

  Update $\mathbf{\Lambda}$:

$$\mathbf{\Lambda}[t+1] = \mathbf{\Lambda}[t] + \mu[t] \left(\mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i[t+1]\mathbf{H}_i[t+1] - \mathbf{G}[t+1] - \mathbf{E}[t+1] \right)$$

  Update μ:

$$\mu[t+1] = \min\left(\rho\mu[t], \mu_{\max}\right)$$

end

parameter. The corresponding partially Augmented Lagrangian for (4) may then be written as

$$\mathcal{L}(\theta) = \lambda \left\| \mathbf{w}_y \odot \hat{\boldsymbol{\epsilon}} \right\|_1 - \frac{1}{2\mu} \left\| \mathbf{\Lambda} \right\|_F^2 + + \frac{\mu}{2} \left\| \mathbf{X} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i - \mathbf{K}\hat{\mathbf{w}} - \hat{\boldsymbol{\epsilon}} + \frac{\mathbf{\Lambda}}{\mu} \right\|_F^2, \quad (20)$$

where $\theta = \left\{ \hat{\mathbf{h}}_i, \hat{\mathbf{w}}, \mathbf{\Lambda} \right\}, i \in \{1,\dots,J\}$. Problem (20) is minimized by employing the ADMM. The algorithm for solving (20) is presented in Algorithm 2. The algorithm terminates when the iterations reach a predefined max value or a convergence criterion is met. The convergence criterion is met when the normalized reconstruction error, i.e., $\left\| \mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t] - \mathbf{K}\hat{\mathbf{w}}[t] - \hat{\boldsymbol{\epsilon}}[t] \right\|_2 / \left\| \mathbf{y} \right\|_2$ is less than a predefined threshold ϵ. The ADMM iteration reads as follows.

Update the primal variables:

For obtaining $\hat{\mathbf{h}}_i, i \in \{1,\dots,J\}$, we need to minimize

$$\hat{\mathbf{h}}_i[t+1] = \underset{\hat{\mathbf{h}}_i[t]}{\operatorname{argmin}} \left\| \mathbf{y} - \sum_{k=1}^{J} \mathbf{F}_k \hat{\mathbf{h}}_k[t] - - \mathbf{K}\hat{\mathbf{w}}[t] - \hat{\boldsymbol{\epsilon}}[t] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right\|_2^2. \quad (21)$$

Problem (21) admits a closed-form solution, which is

$$\hat{\mathbf{h}}_i[t+1] = \mathbf{F}_i^T \Bigg(\mathbf{y} - \sum_{k=1, k\neq i}^{J} \mathbf{F}_k \hat{\mathbf{h}}_k[t] - \mathbf{K}\hat{\mathbf{w}}[t] - \\ - \hat{\boldsymbol{\epsilon}}[t] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \Bigg). \tag{22}$$

For obtaining optimal $\hat{\mathbf{w}}$, we need to minimize

$$\hat{\mathbf{w}}[t+1] = \underset{\hat{\mathbf{w}}[t]}{\operatorname{argmin}} \left\| \mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \\ - \hat{\boldsymbol{\epsilon}}[t] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right\|_2^2 \tag{23}$$

Problem (23) admits a closed-form solution of the form

$$\hat{\mathbf{w}}[t+1] = \mathbf{K}^T \left(\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \hat{\boldsymbol{\epsilon}}[t] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right). \tag{24}$$

For obtaining optimal $\hat{\boldsymbol{\epsilon}}$, we need to minimize

$$\hat{\boldsymbol{\epsilon}}[t] = \underset{\hat{\boldsymbol{\epsilon}}[t]}{\operatorname{argmin}} \left[\lambda \left\| \mathbf{w}_y \odot \hat{\boldsymbol{\epsilon}}[t] \right\|_1 + \frac{\mu[t]}{2} \left\| \mathbf{y} - \\ - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \mathbf{K}\hat{\mathbf{w}}[t+1] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right\|_2^2 \right]. \tag{25}$$

Utilizing the SVT operator, problem (25) admits the following solution

$$\hat{\boldsymbol{\epsilon}}[t+1] = \mathbf{w}_y \odot \mathbf{a} + \overline{\mathbf{w}}_y \odot \Bigg(\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \\ - \mathbf{K}\hat{\mathbf{w}}[t+1] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \Bigg), \tag{26}$$

where

$$\mathbf{a} = \mathcal{S}_{\frac{\lambda}{\mu[t]}} \left(\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \mathbf{K}\hat{\mathbf{w}}[t+1] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right). \tag{27}$$

Update the Lagrange multiplier and parameter μ:

$$\boldsymbol{\lambda}[t+1] = \boldsymbol{\lambda}[t] + \mu[t] \Bigg(\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \\ - \mathbf{K}\hat{\mathbf{w}}[t+1] - \hat{\boldsymbol{\epsilon}}[t+1] \Bigg), \tag{28}$$

$$\mu[t+1] = \min\left(\rho\mu[t], \mu_{\max}\right). \tag{29}$$

III. Experiments

The experimental evaluation of MA-RCA against other state-of-the-art algorithms is carried out via a series of experiments such as: a) noise analysis on synthetic data, b) noise analysis on real facial UV texture maps, c) completion of UV maps with missing values on data captured in "in-the-wild" conditions, d) age-progression on data captured in controlled as well as "in-the-wild" conditions, e) facial expression synthesis on data captured in controlled conditions and comparison with the ground truth corresponding expressions, f) joint *expression* and *illumination* transfer and joint *illumination* and *identity* transfer on data captured in controlled as well as "in-the-wild" conditions.

In order to extract the incomplete face UV maps that were used in our algorithm we have fitted the various databases with a 3DMM. The 3DMM fitting process that was used is the one in [2], which is publicly available. The occlusion masks of the fitted images were extracted by utilizing the corresponding shape and camera parameters.

For the experimental evaluations, databases Multi-PIE [9] and AgeDB [13] were utilized to train MA-RCA. Multi-PIE is a database captured under controlled lab conditions and thus the images do not contain gross-errors attributed to e.g., occlusions. Nevertheless, Multi-PIE is a multi-attribute database, since it contains labels for attributes such as *identity*, *expression* and *illumination*. That renders it suitable to be utilized in MA-RCA to extract bases with respect to e.g., *illumination* that can be then used to reconstruct "in-the-wild" images with various illumination settings (Section III-F). In particular, in the training phase of MA-RCA on Multi-PIE, 90% of the total number of UVs pertaining to distinct identities, expressions and illuminations were utilized. The rest were used for testing.

AgeDB contains images captured under "in-the-wild" conditions (i.e., occlusions, various poses, pixel corruptions are present in the images). Moreover, it is annotated for multiple attributes (i.e., *identity*, *age*) and thus it is suitable for evaluating MA-RCA. AgeDB was split in six distinct age-groups, namely *21-30*, *31-40*, *41-50*, *51-60*, *61-70* and *71-100*. Then, the UVs belonging to each age-group were further split according to the attribute *identity*. In the training phase of MA-RCA, 90% of the total number of UVs were kept to extract the bases with respect to attributes *identity* and *age-groups* and the rest were used for testing.

A. Synthetic noise analysis

In this section we detail the experiments we conducted on synthetic data contaminated by artificial noise of different forms and levels. We report quantitative results and compare the performance of MA-RCA against RJIVE [19], RPCA-L1 [15] and OM-RPCA [16].

In more detail, with the following experiments we aim to investigate the ability of the four component analysis methods to handle data that have been contaminated by two common types of noise. Sparse noise, which is composed of errors that are sparsely supported but of large or unbounded magnitude, is a form of noise that is often present in visual data. Such noise can be the salt & pepper noise, occlusions and registration errors. However, errors of small magnitude, like ambient noise or quantization noise cannot be assumed as sparse errors. For such types of noise it is reasonable to assume that they follow a Gaussian distribution of small variance. Thus, the two types of noise that we consider in our experiments are the gross, sparse, non-Gaussian noise and the Gaussian noise.

Algorithm 2: ADMM solver for problem (20)

Input: $\mathbf{y}$, where $\mathbf{y}$ is the test datum, $\mathbf{F}_i, i \in \{1, \ldots, J\}$ and $\mathbf{G}$, where $\mathbf{F}_i$ and $\mathbf{G}$ are the bases extracted utilizing Algorithm 1, $\mathbf{K}$ is the linear span of $\mathbf{G}$ given by applying the rank-r SVD on $\mathbf{G}$ and $\mathbf{w}_y$, where $\mathbf{w}_y$ is the occlusion mask corresponding the test datum.

Initializations: $t = 0, \epsilon = 1e-8$, maximum number of iterations $t_{\max}, \mu[0] = \frac{25}{\|\mathbf{X}\|}$, $\rho = 1.2$, $\mu_{\max} = 10^7$ and $\lambda = 1/\sqrt{max(F, N)}$.

Output: the reconstructed UV map without the error term, i.e., $\hat{\mathbf{y}} = \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[T] + \mathbf{K}\hat{\mathbf{w}}[T]$

while *not converged* **do**
  for $i = 1 : J$ **do**
    Update $\hat{\mathbf{h}}_i$:
    $$\hat{\mathbf{h}}_i[t+1] = \mathbf{F}_i^T \left(\mathbf{y} - \sum_{k=1, k\neq i}^{J} \mathbf{F}_k \hat{\mathbf{h}}_k[t] - \mathbf{K}\hat{\mathbf{w}}[t] - \hat{\boldsymbol{\epsilon}}[t] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right)$$
  end
  Update $\hat{\mathbf{w}}$:
  $$\hat{\mathbf{w}}[t+1] = \mathbf{K}^T \left(\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_k \hat{\mathbf{h}}_k[t] - \hat{\boldsymbol{\epsilon}}[t] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right)$$
  Update $\hat{\boldsymbol{\epsilon}}$:
  $$\hat{\boldsymbol{\epsilon}}[t+1] = \mathcal{S}_{\frac{\lambda}{\mu[t]}} \left[\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \mathbf{K}\hat{\mathbf{w}}[t+1] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right]$$
  $$\hat{\boldsymbol{\epsilon}}[t+1] = \mathbf{w}_y \odot \hat{\boldsymbol{\epsilon}}[t+1] + \overline{\mathbf{w}}_y \odot \left[\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \mathbf{K}\hat{\mathbf{w}}[t+1] + \frac{\boldsymbol{\lambda}[t]}{\mu[t]} \right]$$
  Update λ:
  $$\boldsymbol{\lambda}[t+1] = \boldsymbol{\lambda}[t] + \mu[t] \left(\mathbf{y} - \sum_{i=1}^{J} \mathbf{F}_i \hat{\mathbf{h}}_i[t+1] - \mathbf{K}\hat{\mathbf{w}}[t+1] - \hat{\boldsymbol{\epsilon}}[t+1] \right)$$
  Update μ:
  $$\mu[t+1] = \min\left(\rho\mu[t], \mu_{\max}\right)$$
end

We conduct experiments for different levels of each form of noise mentioned above.

We generate the data as follows. We create a random matrix $\mathbf{X} \in \mathbf{R}^{m \times n}$ for various dimensions that admits the following decomposition: $\mathbf{X} = \mathbf{F}_*\mathbf{H}_* + \mathbf{G}_* + \mathbf{E}_*$, where $\mathbf{F}_*\mathbf{H}_*$ and $\mathbf{G}_*$ are random matrices with ranks r_1 and r_2, respectively. $\mathbf{E}_*$ corresponds to the noise term. As mentioned previously, we used different noise forms (i.e., sparse non-Gaussian and Gaussian) as well as different noise levels per form. More specifically, we present the results when applying MA-RCA, RJIVE [19], RPCA-L1 [15] and OM-RPCA [16] to extract the error-free components for the following cases:

- *Sparse, non-Gaussian* noise, contaminating $10\%, 25\%$ and 50% of the data matrix entries. The values for the regularizer λ used for the experiments are provided by the formula: $\lambda = 1/\sqrt{max(m, n)}$, following the selection process described in [5]. In particular, for the case of $(m, n) = (500, 500)$, we used $\lambda \approx 0.04$ and for the case of $(m, n) = (1000, 1000)$, we used $\lambda \approx 0.03$. We further empirically show in the supplemental material that this selection formula for λ is indeed the desired one. The convergence threshold ϵ used for this experiment for MA-RCA and RJIVE [19] is $1e-8$. In Table I we report the reconstruction losses where, as evinced, MA-RCA successfully reconstructs the error-free components. In the contrary, RPCA-L1 [15] and OM-RPCA [16] perform poorly compared to RJIVE [19] and MA-RCA, due to the existence of the term $\mathbf{F}_*\mathbf{H}_*$, which is a matrix with identical columns in many parts (e.g., if $\mathbf{F}_*\mathbf{H}_*$ referred to the age-group component, the columns corresponding to the age-group *21-30* would be the same for all of the samples that belong in the specific age-group). The specific structure of $\mathbf{F}_*\mathbf{H}_*$ does not comply with the incoherence property of the low-rank component in the RPCA [5] setting, since $\mathbf{F}_*\mathbf{H}_*$ is a matrix with sparse singular values. As a result, since both RPCA-L1 [15] and OM-RPCA [16] belong in the family of RPCA algorithms, the poor performance under this setting is expected. We empirically validated this fact by conducting synthetic experiments under the setting where the incoherence property for the low-rank component holds. Further details on these experiments can be found in the supplemental material.
- *Gaussian* noise corresponding to zero-mean samples with variances of $0.1, 0.25$ and 0.5. The value for the regularizer λ used for the experiments is very large, i.e., $\lambda =$

$10,000$, since MA-RCA can handle data contaminated by Gaussian noise by vanishing the sparse error term, as explained in Section II-C. The convergence threshold ϵ used for this experiment for MA-RCA and RJIVE [19] is $1e-8$. In Table II we report the reconstruction losses where, as evinced, all the four methods successfully reconstruct the error-free components.

We should note that since MA-RCA constitutes the first multi-attribute learning algorithm, we ran experiments under the single attribute scenario (i.e., only a single term $\mathbf{F}_*\mathbf{H}_*$ is utilized), so that comparisons against the other methods are feasible.

B. Qualitative noise analysis

In this section we detail the experiments we conducted on real data contaminated by artificial noise of different forms and levels. We report qualitative results and compare the performance of MA-RCA against RJIVE [19], RPCA-L1 [15] and OM-RPCA [16].

To pursue this analysis we employ the UV texture data from images of neutral expression from Multi-PIE [9]. Five images of the same person, captured simultaneously in five different viewing angles, $-30°$,$-15°$, $0°$, $15°$ and $30°$ were used to obtain the UV textures. We first fitted a 3DMM on the five images of each person and then generated a corresponding UV map from each image. By using the occlusion free pixels of each of the five UV maps we were able to obtain a full UV texture map that contains no missing values as can also be seen in Fig. 5. We trained all of the four algorithms on UV maps that are completely free of sparse errors (i.e., we only chose subjects that are *not* wearing eye-glasses and also created UV maps without missing values). This is of important essence, as we can then artificially add noise on the error-free UV maps which in turn can be used as ground truth for evaluating the performance of the four methods.

We then applied MA-RCA, RJIVE [19], RPCA-L1 [15] and OM-RPCA [16] to extract the error-free components. Similarly to the quantitative experiments, we present the results for the following cases:

- *Sparse, non-Gaussian* noise with error terms corresponding to samples with $10\%, 25\%$ and 50% non-zero entries. The value for the regularizer λ used for the experiments is provided by the formula: $\lambda = 1/\sqrt{max(m,n)}$, following the selection process described in [5] and the empirical results presented in the supplemental material. In particular, for $n = 765$ UV images and $m = 556266$ the size of a vectorized UV image, $\lambda \approx 0.0015$. The convergence threshold ϵ used for this experiment for MA-RCA and RJIVE [19] is $1e-5$, which is sufficient for the analysis of images. The recovered rank of the individual subspace $\mathbf{G}$ when MA-RCA was applied was 185 for 10% noise, 197 for 25% noise and 216 for 50% noise. These rank values are expected as $\mathbf{G}$ is by definition a low rank matrix which is able to express the individual variations of the data. Also, the low rank of matrix $\mathbf{G}$ is sensible for the data in use as all data are facial images with similar background. As can been seen in Fig. 6, MA-RCA is able to produce better results in recovering the error free images than RPCA-L1 [15] and OM-RPCA [16] in almost all cases. In comparison to RJIVE [19], MA-RCA retains the identity and illumination with greater success.
- *Gaussian* noise with error terms corresponding to zero-mean samples with variances of $0.1, 0.25$ and 0.5. The value for the regularizer λ used for the experiments is very large, i.e., $\lambda = 10,000$, since MA-RCA can handle data contaminated by Gaussian noise by vanishing the sparse error term, as explained previously. The convergence threshold ϵ used for this experiment for MA-RCA and RJIVE [19] is $1e-5$, which is sufficient for the analysis of images. The recovered rank of the individual subspace $\mathbf{G}$ when MA-RCA was applied was 164 for variance 0.1, 182 for variance 0.25 and 214 for variance 0.5. These rank values are expected as in the previous experiment. As anticipated, all four methods are able to recover the error-free images with high accuracy as is also shown in Fig. 7. RPCA-L1 [15] seems to have produced the lowest quality reconstructions.

Similarly to Section III-A, we should note that since MA-RCA constitutes the first multi-attribute learning algorithm, we ran experiments under the single attribute scenario (i.e., only a single term $\mathbf{F}_*\mathbf{H}_*$ is utilized), so that comparisons against the other methods are feasible.

C. Completion of UV maps with missing values

Completion of UV maps with missing values is a very challenging task. Since MA-RCA is inherently able to handle UV maps with missing values, it can be applied on this task. In this experiment, AgeDB "in-the-wild" [13] was utilized.

For the testing phase, a random, incomplete, contaminated with gross but sparse errors UV map which did not belong to the training set was chosen and reconstructed following the process described in Section II-D. As it is evident in Fig. 8 as well as Fig. 9, MA-RCA successfully fills the missing, occluded parts in the original UV map as well as the missing parts in the corresponding 3D facial textures.

D. Age-progression "'in-the-wild"

Age-progression "in-the-wild" entails the task of rendering a facial image of a subject at various ages. It is arguably a very challenging task in Computer Vision, since "in-the-wild" images are captured in uncontrolled conditions (e.g., different illuminations and poses, self-occlusions, etc.). AgeDB [13] was utilized in this experiment, since it is a manually collected "in-the-wild" age database with accurate *age* and *identity* labels and hence the extracted age-groups and identity bases will contain no errors due to incorrect annotations.

During the training phase, MA-RCA was trained under the multi-attribute scenario, incorporating both the knowledge of the *age* as well as the *identity* attributes.

In the testing phase, a random UV map which did not belong to the training set was chosen and reconstructed for various ages following the process described in Section II-D.

Comparisons against other broadly used age-progression methods are provided in Fig. 11 and Fig. 12. More specifically,

TABLE I

QUANTITATIVE RESULTS WHEN COMPARING RECONSTRUCTION LOSSES OF THE ERROR-FREE COMPONENTS AS RETRIEVED BY MA-RCA AGAINST RPCA-L1 [15], OM-RPCA [16] AND RJIVE [19] ON SYNTHETIC DATA OF VARIOUS DIMENSIONS, CONTAMINATED BY SPARSE, NON-GAUSSIAN NOISE, WHERE THE DATA DECOMPOSITION IS $\mathbf{X} = \mathbf{F}_*\mathbf{H}_* + \mathbf{G}_* + \mathbf{E}_*$. IN THE CASE OF RJIVE [19], $\mathbf{FH}$ CORRESPONDS TO THE INDIVIDUAL COMPONENT $\mathbf{A}$ AND $\mathbf{G}$ TO THE JOINT COMPONENT $\mathbf{J}$. IN THE CASE OF RPCA-L1 [15] AND OM-RPCA [16], THE PROJECTION MATRIX $\mathbf{W}$ IS RELATED TO THE LOW RANK COMPONENT $\mathbf{FH} + \mathbf{G}$. ALL OF THE EXPERIMENTS WERE EXECUTED ON AN INTEL I9-7900X 3.30GHZ MACHINE. WE RAN EACH EXPERIMENT 10 TIMES AND REPORT THE AVERAGE EXECUTION TIME FOR EACH ONE.

(m, n, r_1, r_2)	Method	$\frac{\|\mathbf{F}_*\mathbf{H}_* + \mathbf{G}_* - \mathbf{FH} - \mathbf{G}\|_F^2}{\|\mathbf{F}_*\mathbf{H}_* + \mathbf{G}_*\|_F^2}$			Execution time (s)		
	Noise level	10%	25%	50%	10%	25%	50%
(500, 500, 5, 10)	RPCA-L1 [15]	0.182	0.313	2.159	0.99	1.09	0.84
	OM-RPCA [16]	0.181	0.313	2.159	0.36	0.36	0.37
	RJIVE [19]	$3.36e-09$	$3.62e-09$	$6.17e-04$	25.75	26.33	24.38
	MA-RCA	$1.78e-10$	$4.81e-10$	$2.56e-04$	16.25	17.27	16.83
(1000, 1000, 10, 20)	RPCA-L1 [15]	0.183	0.315	2.167	5.04	4.17	4.64
	OM-RPCA [16]	0.183	0.315	2.167	1.72	1.80	1.84
	RJIVE [19]	$1.07e-09$	$2.27e-09$	$8.25e-04$	71.18	71.98	65.11
	MA-RCA	$3.56e-10$	$5.75e-10$	$4.39e-04$	24.85	27.28	28.30

TABLE II

QUANTITATIVE RESULTS WHEN COMPARING RECONSTRUCTION LOSSES OF THE ERROR-FREE COMPONENTS AS RETRIEVED BY MA-RCA AGAINST RPCA-L1 [15], OM-RPCA [16] AND RJIVE [19] ON SYNTHETIC DATA OF VARIOUS DIMENSIONS, CONTAMINATED BY GAUSSIAN NOISE, WHERE THE DATA DECOMPOSITION IS $\mathbf{X} = \mathbf{F}_*\mathbf{H}_* + \mathbf{G}_* + \mathbf{E}_*$. IN THE CASE OF RJIVE [19], $\mathbf{FH}$ CORRESPONDS TO THE INDIVIDUAL COMPONENT $\mathbf{A}$ AND $\mathbf{G}$ TO THE JOINT COMPONENT $\mathbf{J}$. IN THE CASE OF RPCA-L1 [15] AND OM-RPCA [16], THE PROJECTION MATRIX $\mathbf{W}$ IS RELATED TO THE LOW RANK COMPONENT $\mathbf{FH} + \mathbf{G}$. ALL OF THE EXPERIMENTS WERE EXECUTED ON AN INTEL I9-7900X 3.30GHZ MACHINE. WE RAN EACH EXPERIMENT 10 TIMES AND REPORT THE AVERAGE EXECUTION TIME FOR EACH ONE.

(m, n, r_1, r_2)	Method	$\frac{\|\mathbf{F}_*\mathbf{H}_* + \mathbf{G}_* - \mathbf{FH} - \mathbf{G}\|_F^2}{\|\mathbf{F}_*\mathbf{H}_* + \mathbf{G}_*\|_F^2}$			Execution time (s)		
	Variance level	0.10	0.25	0.50	0.10	0.25	0.50
(500, 500, 5, 10)	RPCA-L1 [15]	$6.89e-09$	$4.32e-08$	$1.71e-07$	0.36	0.35	0.35
	OM-RPCA [16]	$6.89e-09$	$4.32e-08$	$1.71e-07$	0.38	0.38	0.36
	RJIVE [19]	$6.89e-09$	$4.32e-08$	$1.71e-07$	31.95	32.34	28.53
	MA-RCA	$6.89e-09$	$4.32e-08$	$1.71e-07$	18.34	18.00	17.65
(1000, 1000, 10, 20)	RPCA-L1 [15]	$1.88e-09$	$1.17e-08$	$4.70e-07$	2.39	1.81	1.83
	OM-RPCA [16]	$1.88e-09$	$1.17e-08$	$4.70e-07$	1.84	1.82	1.84
	RJIVE [19]	$1.88e-09$	$1.17e-08$	$4.70e-07$	142.78	142.03	140.39
	MA-RCA	$1.88e-09$	$1.17e-08$	$4.70e-07$	91.83	86.98	85.04

Fig. 5. UV texture maps extracted by five images of the same person, captured simultaneously from different viewing angles and contained in the multipie database. By combining the five UV maps we are able to generate a full UV texture map.

we compare MA-RCA against Illumination Aware Age Progression (IAAP) [11], Aging with Deep Restricted Boltzmann Machines (ADRBM) [14], Exemplar-based Age Progression (EAP) [22] and RJIVE [19] . Finally, MA-RCA is compared against the state-of-the-art RJIVE [19] in Fig. 10.

E. Facial expression synthesis

In this section, the task of facial expression synthesis under the MA-RCA framework is considered. Facial expression synthesis corresponds to the process of creating artificial images of different expressions from a given facial image.

For this experiment, Multi-PIE [9] was utilized by MA-RCA in order to learn the basis that corresponds to the expression subspace. In particular, the expression subspace of the instantiations *Neutral*, *Disgust*, *Scream*, *Smile*, *Squint* and *Surprise* was extracted.

In the synthesis step, an unseen UV map of expression *Neutral* was chosen and reconstructed for all the learned expressions following the process described in Section II-D. Then, the reconstructed UV maps were projected back on target images of the same identity, illumination and the target expression, included in Multi-Pie [9]. This projection allows a direct qualitative comparison between the ground truth target images and the generated ones. In Fig. 13, examples of synthesized images as well as the target ones are presented. From this figure, it becomes evident that MA-RCA successfully recovers the expression subspace and is capable of generating new instances of different expressions from a given image.

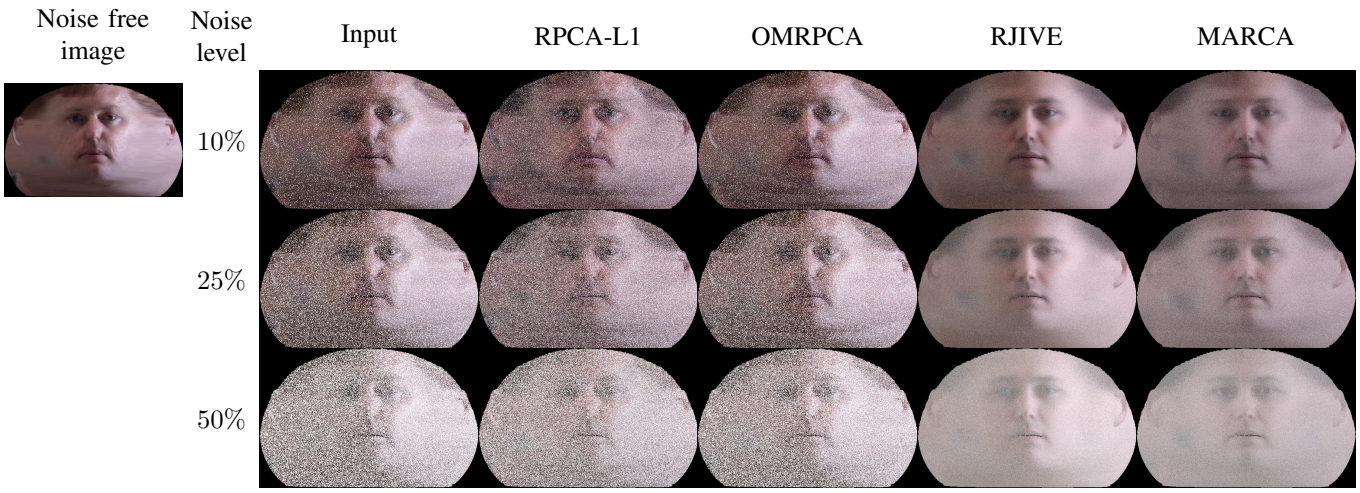

Fig. 6. Reconstructed UV texture maps produced by employing RPCA-L1, OM-RPCA, RJIVE and MA-RCA on data that have been contaminated by different levels of sparse, non-Gaussian noise.

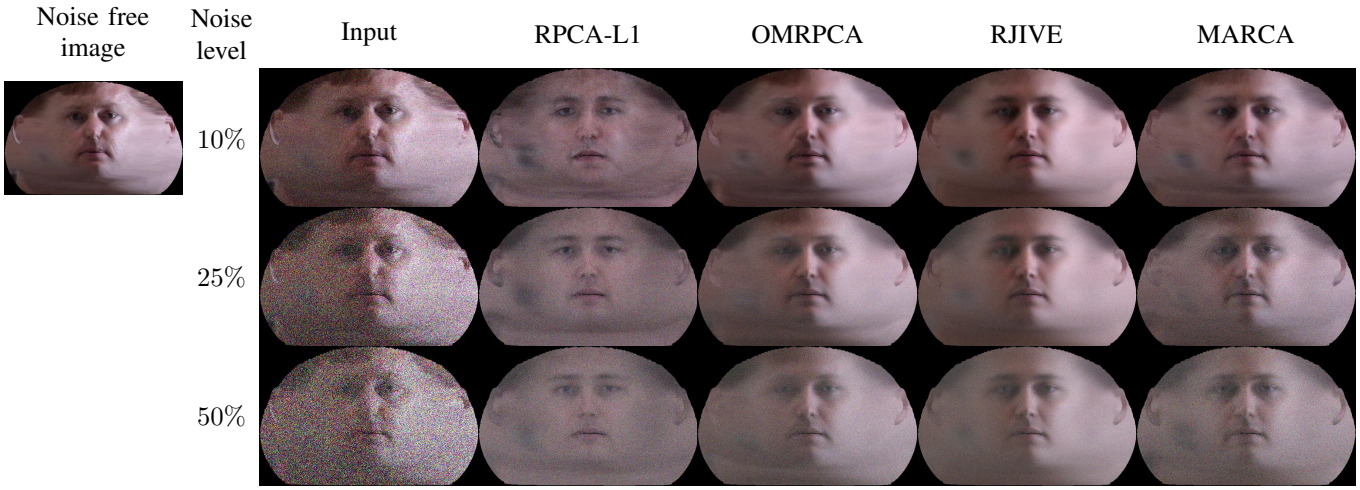

Fig. 7. Reconstructed UV texture maps produced by employing RPCA-L1, OM-RPCA, RJIVE and MA-RCA on data that have been contaminated by different levels of Gaussian noise.

Fig. 8. Examples of UV map completion utilizing MA-RCA. A 3DMM [2] is fitted in the original image and the UV map with missing values and sparse errors is extracted. Finally, the complete UV map is reconstructed utilizing MA-RCA.

F. Multi-attribute transfer "in-the-wild"

In this section, we present a series of experiments under the multi-attribute scenario, i.e., when a test image is reconstructed with more than one attribute transferred at the same time. MA-RCA is the first, to the best of our knowledge, method that can successfully carry out such a task. For this series of experiments, both Multi-PIE [9] and AgeDB [13] are utilized.

In particular, the *illumination* and *expression* bases were extracted from Multi-PIE while the *identity* and *individual* bases were extracted from AgeDB. During the reconstruction of a test image (Section II-D) the bases from Multi-PIE as well as AgeDB are utilized. In Fig. 14, we present how MA-RCA can be utilized to transfer the identity of a particular subject into another one and also transfer the illumination setting at the same time. Additionally, in Fig. 15 we present how expression and illumination can be jointly transferred by employing MA-RCA. For this experiment, we first choose an unseen image of Multi-PIE which depicts a *Neutral* expression and includes no illumination. Then, we are able to synthesize images by combining any instantiation of the learned *illumination* and *expression* bases.

Fig. 9. Examples of 3D texture completion utilizing MA-RCA. Multiple views are visualized to demonstrate the texture completion under various angles.

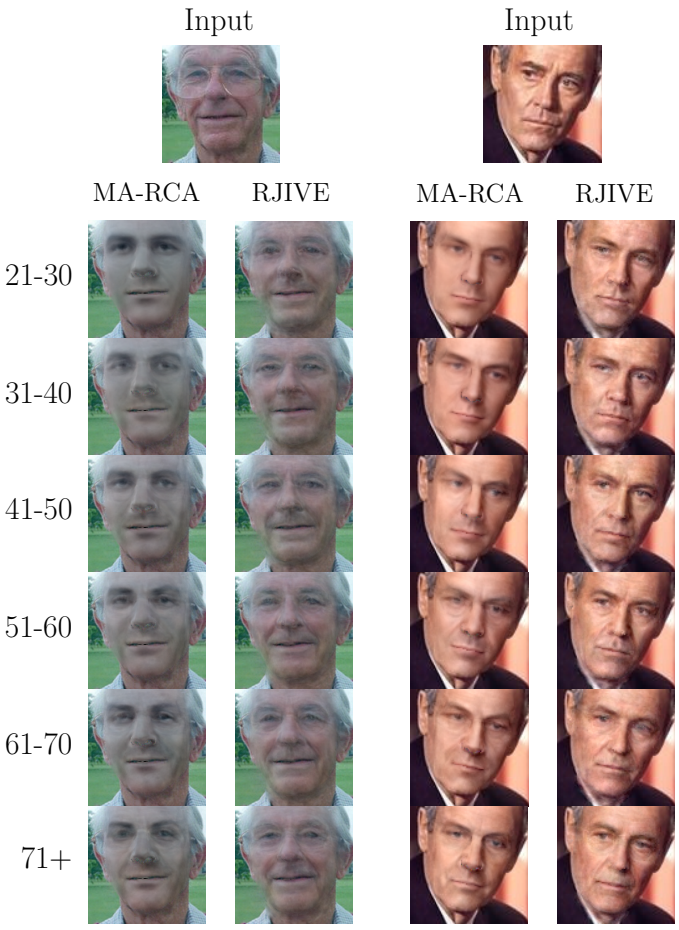

Fig. 10. Comparing MA-RCA against state-of-the-art RJIVE in age-progression "in-the-wild" experiments. As can be seen in both cases, MA-RCA reconstructions are more realistic compared to the reconstructions utilizing RJIVE. More examples can be found in the supplementary material.

IV. Conclusions

With the use of 3D face fitting methods we can generate incomplete facial UV maps of the facial textures. The use of incomplete facial UV maps contaminated with gross errors introduces many challenges and opportunities. In particular, since facial UV map lies in a pose free space, linear component analysis techniques can be applied to learn statistical components for various tasks. In this paper, we propose a novel statistical robust component analysis technique that can tackle the above challenges and at the same time exploit multiple labels of the data at-hand during training. We demonstrate the usefulness of the proposed robust component analysis technique in various tasks including UV map denoisoing, UV map completion on "in-the-wild" data, *illumination*, *expression* and *identity* transfer, as well as aging.

Acknowledgements

Stylianos Moschoglou is funded by an EPSRC DTA studentship from Imperial College London. Evangelos Ververas is funded by a Doctoral Teaching Scholarship from Imperial College London. The work of Stefanos Zafeiriou was partially funded by the EPSRC Project EP/N007743/1 (FACER2VM), as well as by a Google Faculty Award.

References

[1] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.
[2] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou. 3d face morphable models" in-the-wild". *arXiv preprint arXiv:1701.05360*, 2017.
[3] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou. Large scale 3d morphable models.
[4] J. Booth and S. Zafeiriou. Optimal uv spaces for facial morphable model construction. In *IEEE International Conference on Image Processing (ICIP)*, pages 4672–4676, Oct 2014.
[5] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.
[6] T. Cootes and A. Lanitis. The fg-net aging database, 2008.
[7] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.
[8] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.
[9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.
[10] L. A. Jeni, S. Tulyakov, L. Yin, N. Sebe, and J. F. Cohn. The first 3d face alignment in the wild (3dfaw) challenge. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.

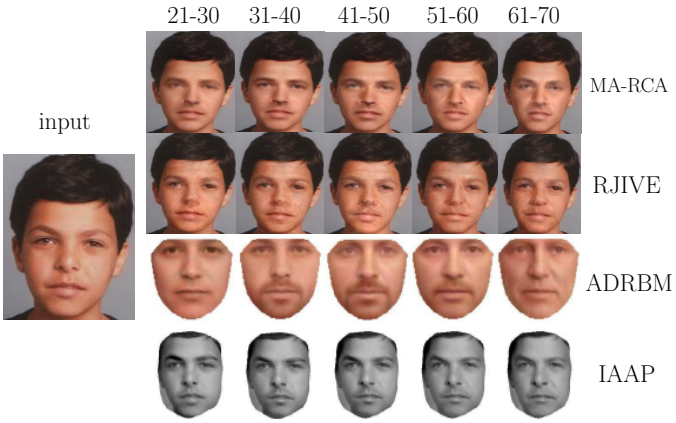

Fig. 11. Age-progression results on FG-Net database [6] produced by IAAP, ADRBM, RJIVE and the proposed MA-RCA method. RJIVE and MA-RCA are trained on AgeDB [13].

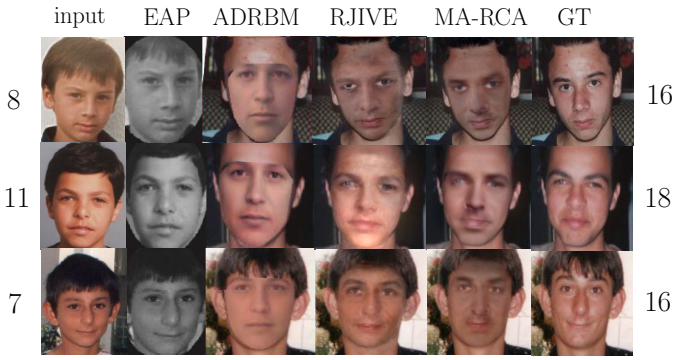

Fig. 12. Age-progression results produced by EAP, ADRBM, RJIVE and the proposed MA-RCA method on images of the FG-Net database [6]. MA-RCA and RJIVE are trained on AgeDB [13].

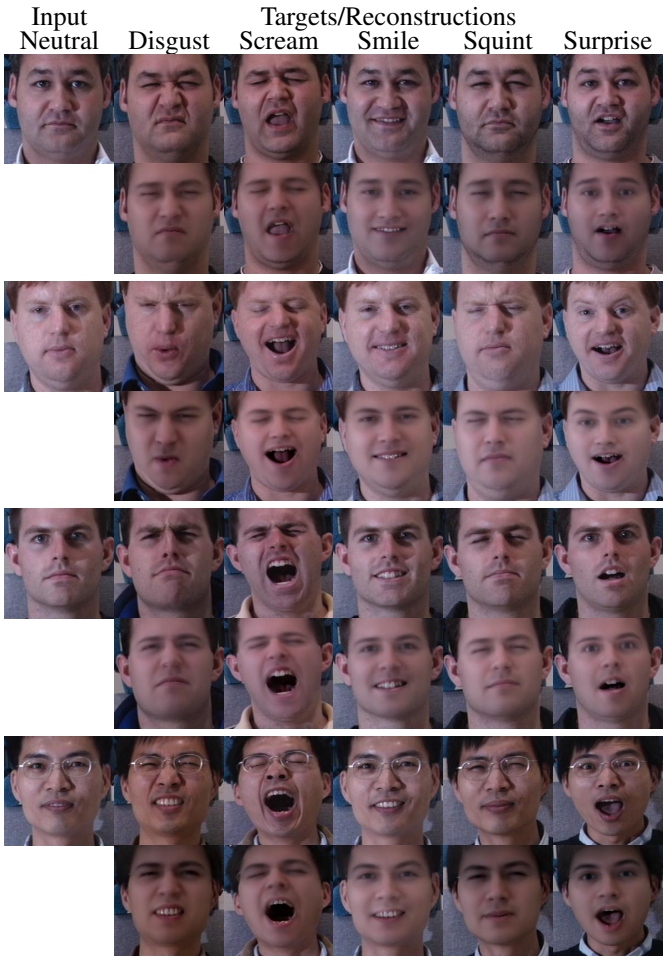

Fig. 13. Facial expression synthesis results on unseen images of Multi-PIE [9]. MA-RCA was employed to extract the expression basis from Multi-PIE [9]. For each subject depicted, the first row includes the input and the target images while the second row includes the reconstructed images of the different expressions.

Fig. 14. Joint transfer of attributes *identity* and *illumination* "in-the-wild". *Illumination* base is learned on Multi-PIE [9] while *identity* base on AgeDB [13]. Final reconstruction of the test image is carried out utilizing both bases.

[11] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz. Illumination-aware age progression. In *Proceedings of the IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, pages 3334–3341, 2014.
[12] H. Kim, M. Zollhöfer, A. Tewari, J. Thies, C. Richardt, and C. Theobalt. Inversefacenet: Deep single-shot inverse face rendering from a single image. *arXiv:1703.10956*, 2017.
[13] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou. Agedb: the first manually collected, in-the-wild age database. 2017.
[14] C. Nhan Duong, K. Luu, K. Gia Quach, and T. D. Bui. Longitudinal face modeling via temporal deep restricted boltzmann machines. In *Proceedings of the IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, pages 5772–5780, 2016.
[15] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang. Robust principal component analysis with non-greedy l1-norm maximization. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 22, page 1433, 2011.
[16] F. Nie, J. Yuan, and H. Huang. Optimal mean robust principal component analysis. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1062–1070, 2014.
[17] Y. Panagakis, M. A. Nicolaou, S. Zafeiriou, and M. Pantic. Robust correlated and individual component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 38(8):1665–1678, 2016.
[18] 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 (PAMI)*, 34(11):2233–2246, 2012.
[19] C. Sagonas, E. Ververas, Y. Panagakis, and S. Zafeiriou. Recovering joint and individual components in facial data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 40(11):2668–2681, Nov 2018.
[20] P. H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.
[21] F. Shang, Y. Liu, J. Cheng, and H. Cheng. Robust principal component analysis with missing data. In *CIKM*, 2014.
[22] C.-T. Shen, W.-H. Lu, S.-W. Shih, and H.-Y. M. Liao. Exemplar-based age progression prediction in children faces. In *Proceedings of the IEEE International Symposium on Multimedia*, pages 123–128. IEEE, 2011.

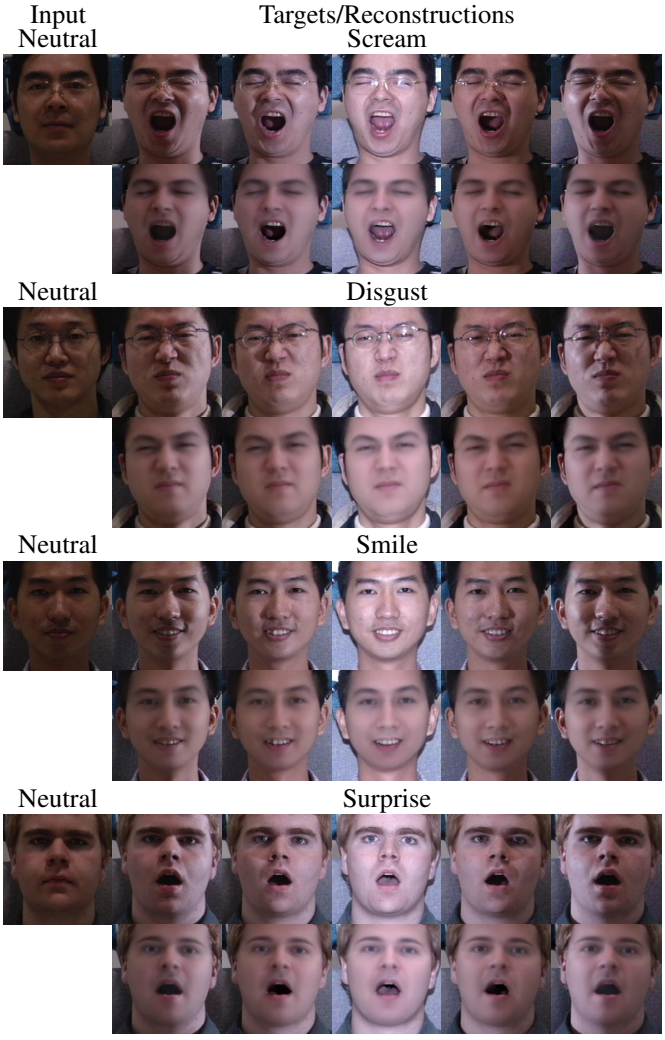

Fig. 15. Joint expression and illumination transfer results on unseen images of Multi-PIE [9]. Expression and illumination bases were learned simultaneously from Multi-PIE [9]. For each subject depicted, the first row includes the expression and illumination free input image as well as the target images. The second row includes the reconstructed images of the target expression and multiple illumination settings.

[23] B. Thompson. Canonical correlation analysis. *Encyclopedia of statistics in behavioral science*, 2005.
[24] S. Zafeiriou, G. G. Chrysos, A. Roussos, E. Ververas, J. Deng, and G. Trigeorgis. The 3d menpo facial landmark tracking challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV-W)*, 2017.
[25] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li. Face alignment across large poses: A 3d solution. In *Proceedings of the IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, 2016.

Stylianos Moschoglou (S'11) is a PhD student at the Department of Computing, Imperial College London, UK, funded by an EPSRC DTA studentship from Imperial College London. Before that, he received the Diploma (MEng) in Electrical and Computer Engineering from Aristotle University of Thessaloniki, Greece, in 2014. Subsequently, Stylianos received the MSc (Distinction) with specialization in Artificial Intelligence from the Department of Computing, Imperial College London, UK, in 2016. Current research interests include statistical machine learning and deep learning with applications to computer vision.

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 Behaviour 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 Senior Lecturer (Associate 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. Dr Panagakis currently serves as the Managing Editor of the Image and Vision Computing Journal. He co-organized the BMVC 2017 conference 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.

Mihalis A. Nicolaou is an Assistant Professor at the Computation-based Science and Technology Research Center at The Cyprus Institute, and a Visiting Academic at Goldsmiths, University of London. Mihalis obtained his PhD from the Department of Computing at Imperial College London under an EPSRC DTA, where he remained as Research Associate and subsequently an Honorary Research Fellow. His research interests span the areas of machine learning, signal processing, and computer vision, focusing on analysis and interpretation of multi-sensory high-dimensional data, often conveyed via visual, auditory, social, and biomedical signals. He has received several awards for his research, including best paper awards at FG '11 and ICASSP '16. Mihalis has co-organized several workshops in his area in top venues such as CVPR, while he has been a guest associate editor at the IEEE Transactions on Affective Computing.

Stefanos Zafeiriou (M09) is currently a Reader in Machine Learning and Computer Vision with the Department of Computing, Imperial College London, London, U.K, a Distinguished Research Fellow with University of Oulu under the Finland Distinguished 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 Presidents Medal for Excellence in Research Supervision in 2016. He is a recipient of many best paper awards including the best student paper award in FG 2018 and ICASSP 2016. In 2018, he received an EPSRC Fellowship and a Google Faculty Award. 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 and Image and Vision Computing Journal. He has been a Guest Editor of over six journal special issues and co-organized over 13 workshops/special sessions on specialized computer vision topics in top venues, such as CVPR/FG/ICCV/ECCV (including three very successful challenges run in ICCV13, ICCV15 and CVPR17 on facial landmark localisation/tracking). He has co-authored over 65 journal papers mainly on novel statistical machine learning methodologies applied to computer vision problems, such as 2D/3D face analysis, deformable object fitting and tracking, shape from shading, and human behavior 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 6500 citations to his work, h-index 41. He was the General Chair of BMVC 2017.

.