Side Information for Face Completion: a Robust PCA Approach

Niannan Xue, *Student Member, IEEE,* Jiankang Deng, *Student Member,IEEE,*
Shiyang Cheng, *Student Member,IEEE,* Yannis Panagakis, *Member,IEEE,*
and Stefanos Zafeiriou, *Member, IEEE*

Abstract—Robust principal component analysis (RPCA) is a powerful method for learning low-rank feature representation of various visual data. However, for certain types as well as significant amount of error corruption, it fails to yield satisfactory results; a drawback that can be alleviated by exploiting domain-dependent prior knowledge or information. In this paper, we propose two models for the RPCA that take into account such side information, even in the presence of missing values. We apply this framework to the task of UV completion which is widely used in pose-invariant face recognition. Moreover, we construct a generative adversarial network (GAN) to extract side information as well as subspaces. These subspaces not only assist in the recovery but also speed up the process in case of large-scale data. We quantitatively and qualitatively evaluate the proposed approaches through both synthetic data and eight real-world datasets to verify their effectiveness.

Index Terms—RPCA, GAN, side information, UV completion, face recognition, in the wild.

1 Introduction

UV space embeds the manifold of a 3D face into a 2D contiguous atlas. Contiguous UV spaces are natural products of many 3D scanning devices and are often used by 3D Morphable Model (3DMM) construction [1], [2], [3]. Although UV space by nature cannot be constructed from an arbitrary 2D image, a UV map can still be obtained by fitting a 3DMM to the image and sampling the corresponding texture [4]. We illustrate this procedure in Figure 1. Unfortunately, due to self-occlusion of the face, those UV maps are often incomplete and lack facial parts that are informative. Once completed, this UV map, combined with the corresponding 3D face, is extremely useful, as it can be used to synthesise 2D faces of arbitrary poses. Afterwards, we can probe image pairs of similar poses to improve recognition performance [5]. Hence, the success of pose-invariant face recognition relies on the quality of UV map completion.

Recovering UV maps from a sequence of related facial frames is a challenging task because self-occlusion at large poses leads to incomplete and missing data. Meanwhile, the imperfection in fitting leads to regional errors. We adapt the approach of robust principal component analysis (RPCA) with missing data [6] to address this difficult problem. In other words, we operate directly on the images themselves rather than on their labels [7]. Principal Component Pursuit (PCP) as proposed in [8], [9] and its variants e.g., [10], [11], [12], [13], [14], [15], [16] are popular algorithms to solve RPCA. PCP employs the nuclear norm and the l_1-norm (convex surrogates of the rank and sparsity constraints, respectively) in order to approximate the original l_0-norm regularised rank minimisation problem. Unavoidably, PCP operates in an isolated manner where domain-dependent prior knowledge [17], i.e., side information [18], is always ignored. Moreover, real-world visual data rarely satisfies the stringent assumptions imposed by PCP for exact recovery [19]. These call for a more powerful framework that can assimilate useful priors to alleviate the degenerate or sub-optimal solutions of PCP.

It has already been shown that side information is propitious in the context of matrix completion [20], [21] and compressed sensing [22]. Recently, *noiseless* features have been capitalised on in the PCP framework [23], [24], [25], [26]. In particular, an error-free orthogonal column space was used to drive a person-specific facial deformable model [24]. And such features can also remove dependency on the row-coherence which is beneficial in the case of a union of multiple subspaces [25], [26], [27], [28]. More generally, Chiang et al. [23] used both a column and a row space to recover only the weights of their interaction in a simpler problem. The main hindrance to the success of these methods is the need for a set of clean, noise-free data samples in order to determine the column and/or row spaces of the low-rank component. But there are no prescribed way to find them in practice.

On a separate note, rapid advances in neural networks for image inpainting offer an agglomeration of useful priors. Pathak et al. [29] proposed to use context encoders with a reconstruction and an adversarial loss to generate contents for the missing regions that comply with the neighbourhood. Yang et al. [30] further improved inpainting with a multi-scale neural patch synthesis method. This approach is based on a joint optimisation of image content and texture constraints, which not only preserves contextual structures but also produces fine details. Li et al. [31] combined a

- *N. Xue, J. Deng, S. Cheng, Y. Panagakis and S. Zafeiriou are with the Department of Computing, Imperial College London, UK. Corresponding author: Jiankang Deng, E-mail: j.deng16@imperial.ac.uk*
- *S. Zafeiriou is also with Center for Machine Vision and Signal Analysis, University of Oulu, Finland. J. Deng and S. Zafeiriou are also with Facesoft.io.*

Manuscript received on September 20, 2018; revised on January 7, 2019; accepted on February 23, 2019.

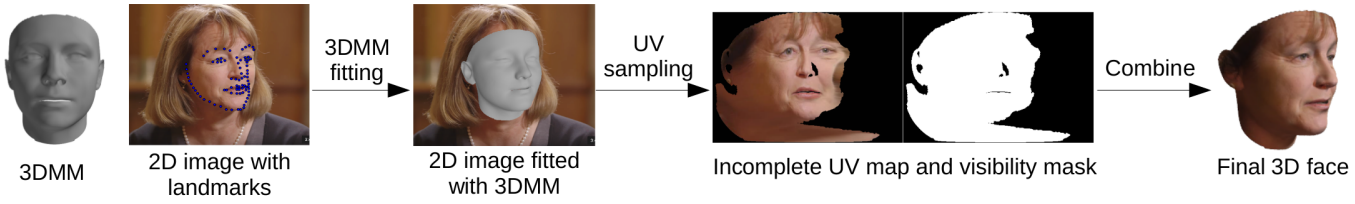


Fig. 1. The procedure of getting the UV map from an arbitrary 2D image.

reconstruction loss, two adversarial losses, and a semantic parsing loss to ensure genuineness and consistency of local-global contents. These methods are by no means definitive for the following reasons: (a) their masks are artificial and do not have semantic correspondence with a 3D face; (b) they do not allow missing regions to be over 50% which is commonplace in our case.

This paper is based on our preliminary work [32] but has been extended to 1) the problem of UV completion and 2) to incorporate side information provided by generative adversarial networks. As such, we have extended PCP to take advantage of *noisy* prior information aiming to realise better UV map reconstruction. We then perform pose-invariant face recognition experiments using the completed UV maps. Experimental results indicate the superiority of our framework. The overall workflow is explicated in Figure 2. Our contributions are summarised as follows:

- A novel convex program is proposed to use side information, which is a noisy approximation of the low-rank component, within the PCP framework. The proposed method is able to handle missing values while the developed optimisation algorithm has convergence guarantees.
- Furthermore, we extend our proposed PCP model using side information to exploit prior knowledge regarding the column and row spaces of the low-rank component in a more general algorithmic framework.
- In the case of UV completion, we suggest the use of generative adversarial networks to provide subspace features and side information, resulting in a seamless integration of deep learning into the robust PCA framework.
- We demonstrate the applicability and effectiveness of the proposed approaches on synthetic data as well as on facial image denoising, UV texture completion and pose-invariant face recognition experiments with both quantitative and qualitative evaluation.

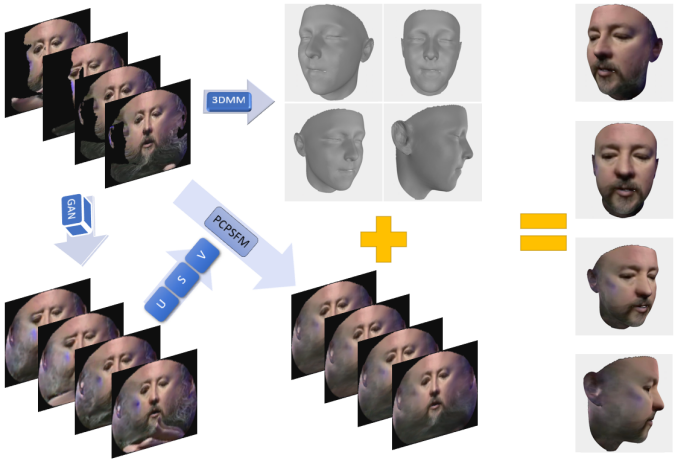


Fig. 2. Given an input sequence of incomplete UV maps, we extract the shape using 3DMM and perform preliminary completion using GAN. With the left subspace and side information provided by GAN, we then carry out PCPSFM to produce more refined completion results. After that, we attach the completed UV texture to the shape creating images at various poses for face recognition.

The remainder of this paper is organised as follows. We discuss relevant literature in Section 2, while the proposed robust principal component analysis using side information with missing values (PCPSM) along with its extension that incorporates features (PCPSFM) is presented in Section 3. In Section 4, we first evaluate our proposed algorithms on synthetic and real-world data. Then we introduce GAN as a source of features and side information for the subject of UV completion. Finally, face recognition experiments are presented in the last subsection.

Notations Lowercase letters denote scalars and uppercase letters denote matrices, unless otherwise stated. For norms of matrix $\mathbf{A}$, $\|\mathbf{A}\|_F$ is the Frobenius norm; $\|\mathbf{A}\|_*$ is the nuclear norm; and $\|\mathbf{A}\|_1$ is the sum of absolute values of all matrix entries. Moreover, $\langle \mathbf{A}, \mathbf{B} \rangle$ represents $\mathrm{tr}(\mathbf{A}^T\mathbf{B})$ for real matrices $\mathbf{A}, \mathbf{B}$. Additionally, $\mathbf{A} \circ \mathbf{B}$ symbolises element-wise multiplication of two matrices of the same dimension.

2 Related Work

We discuss two different lines of research, namely low-rank recovery as well as image completion.

2.1 Robust principal component analysis

Suppose that there is a matrix $\mathbf{L}_0 \in \mathbb{R}^{n_1 \times n_2}$ with rank $r \ll \min(n_1, n_2)$ and a sparse matrix $\mathbf{E}_0 \in \mathbb{R}^{n_1 \times n_2}$ with entries of arbitrary magnitude. If we are provided with the observation matrix $\mathbf{X} = \mathbf{L}_0 + \mathbf{E}_0$, RPCA aims to recover them by solving the following objective:

$$\min_{\mathbf{L},\mathbf{E}} \mathrm{rank}(\mathbf{L}) + \lambda \|\mathbf{E}\|_0 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{L} + \mathbf{E}, \tag{1}$$

where λ is a regularisation parameter. However, (1) cannot be readily solved because it is NP-hard. PCP instead solves the following convex surrogate:

$$\min_{\mathbf{L},\mathbf{E}} \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{L} + \mathbf{E}, \tag{2}$$

which, under mild conditions, is equivalent to (1). There exist many efficient solvers for (2) and its applications include background modelling from surveillance video and removing shadows and specularities from face images.

One of the first methods for incorporating dictionary was proposed in the context of subspace clustering [25], [26]. The LRR algorithm assumes that we have available an orthogonal column space $\mathbf{U} \in \mathbb{R}^{n_1 \times d_1}$, where $d_1 \leq n_1$, and optimises the following:

$$\min_{\mathbf{K},\mathbf{E}} \quad \|\mathbf{K}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{U}\mathbf{K} + \mathbf{E}. \tag{3}$$

Given an orthonormal statistical prior of facial images, LRR can be used to construct person-specific deformable models from erroneous initialisations [24].

A generalisation of the above was proposed as Principal Component Pursuit with Features (PCPF) [23] where further row spaces $\mathbf{V} \in \mathbb{R}^{n_2 \times d_2}$, $d_2 \leq n_2$, were assumed to be available with the following objective:

$$\min_{\mathbf{H},\mathbf{E}} \quad \|\mathbf{H}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{U}\mathbf{H}\mathbf{V}^T + \mathbf{E}. \tag{4}$$

There is a stronger equivalence relation between (4) and (1) than (2). The main drawback of the above mentioned models is that features need to be accurate and noiseless, which is not trivial to fulfil in practical scenarios.

In the case of missing data, robust matrix recovery methods [6], [33] enhance PCP to deal with occlusions:

$$\min_{\mathbf{L},\mathbf{E}} \|\mathbf{L}\|_* + \lambda\|\mathbf{E} \circ \mathbf{W}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{L} + \mathbf{E}, \tag{5}$$

where $\mathbf{W}$ is the matrix of binary occlusion masks. Its Jacobi-type update schemes can be implemented in parallel and hence are attractive for solving large-scale problems. Disgruntled at the unrealistic uniform sampling assumption for missing entries, Liu et al [26] set out to use the isomeric condition hypothesis to tackle irregular and deterministic missing data.

2.2 Image completion neural networks

Recent advances in convolutional neural networks (CNN) also show great promises in visual feature learning. Context encoders (CE) [29] use an encoder-decoder pipeline where the encoder takes an input image with missing regions producing a latent feature representation and the decoder takes the feature representation generating the missing image content. CE uses a joint loss function:

$$\mathcal{L} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{adv}\mathcal{L}_{adv}, \tag{6}$$

where $\mathcal{L}_{rec}$ is the reconstruction loss and $\mathcal{L}_{adv}$ is the adversarial loss. The reconstruction loss is given by:

$$\mathcal{L}_{rec}(\mathbf{x}) = \|\mathbf{w} \circ (\mathbf{x} - F((\mathbf{1} - \mathbf{w}) \circ x)\|_2^2, \tag{7}$$

where $\mathbf{w}$ is a binary mask, $\mathbf{x}$ is an example image and CE produces an output $F(\mathbf{x})$. The adversarial loss is based on Generative Adversarial Networks (GAN). GAN learns both a generative model G_i from noise distribution $\mathcal{Z}$ to data distribution $\mathcal{X}$ and a discriminative model D_i by the following objective:

$$\mathcal{L}_{a_i} = \min_{G_i} \max_{D_i} \mathbb{E}_{\mathbf{x}\in\mathcal{X}}[\log(D_i(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\in\mathcal{Z}}[\log(1 - D_i(G_i(\mathbf{z})))]. \tag{8}$$

For CE, the adversarial loss is modified to

$$\mathcal{L}_{adv} = \max_{D} \mathbb{E}_{\mathbf{x}\in\mathcal{X}}[\log(D(\mathbf{x})) + \log(1 - D(F((\mathbf{1} - \mathbf{w}) \circ \mathbf{x})))]. \tag{9}$$

Generative face completion [31] uses two discriminators instead with the following objective

$$\mathcal{L} = \mathcal{L}_p + \lambda_1\mathcal{L}_{a_1} + \lambda_2\mathcal{L}_{a_2}, \tag{10}$$

where $\mathcal{L}_p$ is a parsing loss of pixel-wise softmax between the estimated UV texture $I_{i,j}$ and the ground truth texture $I^*_{i,j}$ of width W and height H

$$L_p = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \left| I_{i,j} - I^*_{i,j} \right|. \tag{11}$$

Patch synthesis [30] optimises a loss function of three terms: the holistic content term, the local texture term and the TV-loss term. The content constraint penalises the l_2 difference between the optimisation result and the previous content prediction

$$l_c = \|\mathbf{w} \circ (\mathbf{x} - \mathbf{x}_i)\|_2^2, \tag{12}$$

where $\mathbf{x}_i$ is the optimisation result from the last iteration at a coarser scale. The texture constraint penalises the texture appearance across the hole,

$$l_t = \frac{1}{|\mathbf{w}^\phi|} \sum_{i \in \mathbf{w}^\phi} \|P_i \circ \phi(\mathbf{x}) - P_{nn(i)} \circ \phi(\mathbf{x})\|_2^2, \tag{13}$$

where $\mathbf{w}^\phi$ is the corresponding mask in the VGG-19 feature map $\phi(\mathbf{x})$, $|\mathbf{w}^\phi|$ denotes the number of patches sampled in $\mathbf{w}^\phi$, P_i is the local neural patch at location i, and $nn(i)$ is the nearest neighbor of i. Last, the TV loss encourages smoothness:

$$l_{TV} = \sum_{i,j \in \mathbf{w}^\phi} ((\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j})^2 + (\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j})^2). \tag{14}$$

3 Robust Principal Component Analysis Using Side Information

In this section, we propose models of RPCA using side information. In particular, we incorporate side information into PCP by using the trace distance of the difference between the low-rank component and the noisy estimate, which can be seen as a generalisation of compressed sensing with prior information where l_1 norm has been used to minimise the distance between the target signal and side information [22].

3.1 The PCPSM and PCPSFM models

Assuming that a noisy estimate of the low-rank component of the data $\mathbf{S} \in \mathbb{R}^{n_1 \times n_2}$ is available, we propose the following model of PCP using side information with missing values (PCPSM):

$$\begin{aligned} \min_{\mathbf{L},\mathbf{E}} \quad & \|\mathbf{L}\|_* + \alpha\|\mathbf{L} - \mathbf{S}\|_* + \lambda\|\mathbf{W} \circ \mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{L} + \mathbf{E}, \end{aligned} \tag{15}$$

where $\alpha > 0, \lambda > 0$ are parameters that weigh the effects of side information and noise sparsity.

The proposed PCPSM can be revamped to generalise the previous attempt of PCPF by the following objective of PCP

using side information with features and missing values (PCPSFM):

$$\begin{aligned} \min_{\mathbf{H},\mathbf{E}} \quad & \|\mathbf{H}\|_* + \alpha\|\mathbf{H}-\mathbf{D}\|_* + \lambda\|\mathbf{W}\circ\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{U}\mathbf{H}\mathbf{V}^T + \mathbf{E}, \quad \mathbf{D} = \mathbf{U}^T\mathbf{S}\mathbf{V}, \end{aligned} \tag{16}$$

where $\mathbf{H} \in \mathbb{R}^{d_1\times d_2}, \mathbf{D} \in \mathbb{R}^{d_1\times d_2}$ are bilinear mappings for the recovered low-rank matrix $\mathbf{L}$ and side information $\mathbf{S}$ respectively. Note that the low-rank matrix $\mathbf{L}$ is recovered from the optimal solution $(\mathbf{H}^*, \mathbf{E}^*)$ to objective (16) via $\mathbf{L} = \mathbf{U}\mathbf{H}^*\mathbf{V}^T$. If side information $\mathbf{S}$ is not available, PCPSFM reduces to PCPF with missing values by setting α to zero. If the features $\mathbf{U}, \mathbf{V}$ are not present either, PCP with missing values can be restored by fixing both of them at identity. However, when only the side information $\mathbf{S}$ is accessible, objective (16) is transformed back into PCPSM.

3.2 The algorithm

If we substitute $\mathbf{B}$ for $\mathbf{H}-\mathbf{D}$ and orthogonalise $\mathbf{U}$ and $\mathbf{V}$, the optimisation problem (16) is identical to the following convex but non-smooth problem:

$$\begin{aligned} \min_{\mathbf{H},\mathbf{E}} \quad & \|\mathbf{H}\|_* + \alpha\|\mathbf{B}\|_* + \lambda\|\mathbf{W}\circ\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{U}\mathbf{H}\mathbf{V}^T + \mathbf{E}, \quad \mathbf{B} = \mathbf{H} - \mathbf{U}^T\mathbf{S}\mathbf{V}, \end{aligned} \tag{17}$$

which is amenable to the multi-block alternating direction method of multipliers (ADMM).

The corresponding augmented Lagrangian of (17) is:

$$\begin{aligned} l(\mathbf{H},\mathbf{B},\mathbf{E},\mathbf{Z},\mathbf{N}) &= \|\mathbf{H}\|_* + \alpha\|\mathbf{B}\|_* + \lambda\|\mathbf{W}\circ\mathbf{E}\|_1 \\ &+ \langle\mathbf{Z}, \mathbf{X}-\mathbf{E}-\mathbf{U}\mathbf{H}\mathbf{V}^T\rangle + \frac{\mu}{2}\|\mathbf{X}-\mathbf{E}-\mathbf{U}\mathbf{H}\mathbf{V}^T\|_F^2 \\ &+ \langle\mathbf{N}, \mathbf{H}-\mathbf{B}-\mathbf{U}^T\mathbf{S}\mathbf{V}\rangle + \frac{\mu}{2}\|\mathbf{H}-\mathbf{B}-\mathbf{U}^T\mathbf{S}\mathbf{V}\|_F^2, \end{aligned} \tag{18}$$

where $\mathbf{Z} \in \mathbb{R}^{n_1\times n_2}$ and $\mathbf{N} \in \mathbb{R}^{d_1\times d_2}$ are Lagrange multipliers and μ is the learning rate.

The ADMM operates by carrying out repeated cycles of updates till convergence. During each cycle, $\mathbf{H}, \mathbf{B}, \mathbf{E}$ are updated serially by minimising (18) with other variables fixed. Afterwards, Lagrange multipliers $\mathbf{Z}, \mathbf{N}$ are updated at the end of each iteration. Direct solutions to the single variable minimisation subproblems rely on the shrinkage and the singular value thresholding operators [8]. Let $\mathcal{S}_\tau(a) \equiv \text{sgn}(a)\max(|a|-\tau, 0)$ serve as the shrinkage operator, which naturally extends to matrices, $\mathcal{S}_\tau(\mathbf{A})$, by applying it to matrix $\mathbf{A}$ element-wise. Similarly, let $\mathcal{D}_\tau(\mathbf{A}) \equiv \mathbf{M}\mathcal{S}_\tau(\mathbf{\Sigma})\mathbf{Y}^T$ be the singular value thresholding operator on real matrix $\mathbf{A}$, with $\mathbf{A} = \mathbf{M}\mathbf{\Sigma}\mathbf{Y}^T$ being the singular value decomposition (SVD) of $\mathbf{A}$.

Minimising (18) w.r.t. $\mathbf{H}$ at fixed $\mathbf{B}, \mathbf{E}, \mathbf{Z}, \mathbf{N}$ is equivalent to the following:

$$\arg\min_{\mathbf{H}} \ \|\mathbf{H}\|_* + \mu\|\mathbf{P} - \mathbf{U}\mathbf{H}\mathbf{V}^T\|_F^2, \tag{19}$$

where $\mathbf{P} = \frac{1}{2}(\mathbf{X}-\mathbf{E}+\frac{1}{\mu}\mathbf{Z}+\mathbf{U}(\mathbf{B}+\mathbf{U}^T\mathbf{S}\mathbf{V}-\frac{1}{\mu}\mathbf{N})\mathbf{V}^T)$. Its solution is shown to be $\mathbf{U}^T\mathcal{D}_{\frac{1}{2\mu}}(\mathbf{P})\mathbf{V}$. Furthermore, for $\mathbf{B}$,

$$\arg\min_{\mathbf{B}} \ l = \arg\min_{\mathbf{B}} \ \alpha\|\mathbf{B}\|_* + \frac{\mu}{2}\|\mathbf{Q}-\mathbf{B}\|_F^2, \tag{20}$$

where $\mathbf{Q} = \mathbf{H} - \mathbf{U}^T\mathbf{S}\mathbf{V} + \frac{1}{\mu}\mathbf{N}$, whose update rule is $\mathcal{D}_{\frac{\alpha}{\mu}}(\mathbf{Q})$, and for $\mathbf{E}$,

$$\arg\min_{\mathbf{E}} \ l = \arg\min_{\mathbf{E}} \ \lambda\|\mathbf{W}\circ\mathbf{E}\|_1 + \frac{\mu}{2}\|\mathbf{R}-\mathbf{E}\|_F^2, \tag{21}$$

Algorithm 1 ADMM solver for PCPSFM

Input: Observation $\mathbf{X}$, mask $\mathbf{W}$, side information $\mathbf{S}$, features $\mathbf{U}, \mathbf{V}$, parameters $\alpha, \lambda > 0$, scaling ratio $\beta > 1$.

1: **Initialize:** $\mathbf{Z} = 0$, $\mathbf{N} = \mathbf{B} = \mathbf{H} = 0$, $\beta = \frac{1}{\|\mathbf{X}\|_2}$.
2: **while** not converged **do**
3: $\quad\mathbf{E} = \mathcal{S}_{\lambda\mu^{-1}}(\mathbf{X} - \mathbf{U}\mathbf{H}\mathbf{V}^T + \frac{1}{\mu}\mathbf{Z})\circ\mathbf{W} + (\mathbf{X} - \mathbf{U}\mathbf{H}\mathbf{V}^T + \frac{1}{\mu}\mathbf{Z})\circ(\mathbf{1}-\mathbf{W})$
4: $\quad\mathbf{H} = \mathbf{U}^T\mathcal{D}_{\frac{1}{2\mu}}(\frac{1}{2}(\mathbf{X} - \mathbf{E} + \frac{1}{\mu}\mathbf{Z} + \mathbf{U}(\mathbf{B} + \mathbf{U}^T\mathbf{S}\mathbf{V} - \frac{1}{\mu}\mathbf{N})\mathbf{V}^T))\mathbf{V}$
5: $\quad\mathbf{B} = \mathcal{D}_{\alpha\mu^{-1}}(\mathbf{H} - \mathbf{U}^T\mathbf{S}\mathbf{V} + \frac{1}{\mu}\mathbf{N})$
6: $\quad\mathbf{Z} = \mathbf{Z} + \mu(\mathbf{X} - \mathbf{E} - \mathbf{U}\mathbf{H}\mathbf{V}^T)$
7: $\quad\mathbf{N} = \mathbf{N} + \mu(\mathbf{H} - \mathbf{B} - \mathbf{U}^T\mathbf{S}\mathbf{V})$
8: $\quad\mu = \mu \times \beta$
9: **end while**

Return: $\mathbf{L} = \mathbf{U}\mathbf{H}\mathbf{V}^T, \mathbf{E}$

where $\mathbf{R} = \mathbf{X} - \mathbf{U}\mathbf{H}\mathbf{V}^T + \frac{1}{\mu}\mathbf{Z}$ with a closed-form solution $\mathcal{S}_{\lambda\mu^{-1}}(\mathbf{R})\circ\mathbf{W} + \mathbf{R}\circ(\mathbf{1}-\mathbf{W})$. Finally, Lagrange multipliers are updated as usual:

$$\mathbf{Z} = \mathbf{Z} + \mu(\mathbf{X} - \mathbf{E} - \mathbf{U}\mathbf{H}\mathbf{V}^T), \tag{22}$$

$$\mathbf{N} = \mathbf{N} + \mu(\mathbf{H} - \mathbf{B} - \mathbf{U}^T\mathbf{S}\mathbf{V}). \tag{23}$$

The overall algorithm is summarised in Algorithm 1.

3.3 Complexity and convergence

Orthogonalisation of the features $\mathbf{U}, \mathbf{V}$ via the Gram-Schmidt process has an operation count of $O(n_1 d_1^2)$ and $O(n_2 d_2^2)$ respectively. The $\mathbf{H}$ update in Step 4 is the most costly step of each iteration in Algorithm 1. Specifically, the SVD required in the singular value thresholding action dominates with $O(\min(n_1 n_2^2, n_1^2 n_2))$ complexity. Note that this complexity is shared by both of our proposed PCPSM and PCPSFM algorithms, as well as exsiting PCP and LRR algorithms.

A direct extension of the ADMM has been applied to our 3-block separable convex objective. Its global convergence is proved in **Theorem 1**. We have also used the fast continuation technique already applied to the matrix completion problem [34] to increase μ incrementally for accelerated superlinear performance [35]. The cold start initialisation strategies for variables $\mathbf{H}, \mathbf{B}$ and Lagrange multipliers $\mathbf{Z}, \mathbf{N}$ are described in [36]. Besides, we have scheduled $\mathbf{E}$ to be updated first and taken the initial learning rate μ as suggested in [37]. As for stopping criteria, we have employed the Karush-Kuhn-Tucker (KKT) feasibility conditions. Namely, within a maximum number of 1000 iterations, when the maximum of $\|\mathbf{X}-\mathbf{E}_k-\mathbf{U}\mathbf{H}_k\mathbf{V}^T\|_F/\|\mathbf{X}\|_F$ and $\|\mathbf{H}_k - \mathbf{B}_k - \mathbf{U}^T\mathbf{S}\mathbf{V}\|_F/\|\mathbf{X}\|_F$ dwindles from a predefined threshold ϵ, the algorithm is terminated, where k signifies values at the k^{th} iteration.

Theorem 1. *Let the iterative squence $\{(\mathbf{E}^k, \mathbf{H}^k, \mathbf{B}^k, \mathbf{Z}^k, \mathbf{N}^k)\}$ be generated by the direct extension of ADMM, Algorithm 1, then the sequence $\{(\mathbf{E}^k, \mathbf{H}^k, \mathbf{B}^k, \mathbf{Z}^k, \mathbf{N}^k)$ converges to a Karush-Kuhn-Tucher (KKT) point in the fully observed case.*

Proof. We first show that function $\theta_3(x_3) = \|E\|_1$ is sub-strong monotonic. From [8], we know that $(x_1^*, x_2^*, x_3^*, \lambda^*) =$

$(\mathbf{H}_0, \mathbf{E}_0, \mathbf{B}_0, \mathbf{Z}_0)$ is a KKT point, where $\mathbf{H}_0 = \mathbf{U}^T\mathbf{L}_0\mathbf{V}$, $\mathbf{B}_0 = \mathbf{H}_0 - \mathbf{U}^T\mathbf{S}\mathbf{V}$, $\mathbf{Z}_{0ij} = \lambda[\text{sgn}(\mathbf{E}_0)]_{ij}$, if $(i,j) \in \Omega$ and $|\mathbf{Z}_{0ij}| < \lambda$, otherwise. Since $\theta_3(x_3)$ is convex, by definition, we have

$$\theta_3(x_3^*) \geq \theta_3(x_3) + \langle y_3, x_3^* - x_3\rangle, \quad \forall x_3 \text{ and } \forall y_3 \in \partial\theta_3(x_3). \tag{24}$$

Since A_3 is identity in (17), we have

$$\begin{aligned} &\theta_3(x_3) - \theta_3(x_3^*) + \langle A_3^T\lambda^*, x_3^* - x_3\rangle \\ =&\lambda\|\mathbf{E}\|_1 - \lambda\|\mathbf{E}_0\|_1 + \langle \mathbf{Z}_0, \mathbf{E}_0\rangle - \langle \mathbf{Z}_0, \mathbf{E}\rangle, \\ =&\lambda\|\mathbf{E}\|_1 - \langle \mathbf{Z}_0, \mathbf{E}\rangle \\ \geq&0, \end{aligned} \tag{25}$$

where the third line follows from $\mathbf{Z}_{0ij} = \lambda[\text{sgn}(\mathbf{E}_0)]_{ij}$ when $(i,j) \in \Omega$ and $\mathbf{E}_{0ij} = 0$ when $(i,j) \notin \Omega$, and the fourth line follows from $|\mathbf{Z}_{0ij}| \leq \lambda$, $|\mathbf{Z}_{0ij}\mathbf{E}_{ij}| \leq |\mathbf{Z}_{0ij}||\mathbf{E}_{ij}|$ and $\|\mathbf{E}\|_1 = \sum_{i,j}|\mathbf{E}_{ij}|$. As $\mathbf{E}$ is bounded, there always exists $\mu > 0$ such that

$$\lambda\|\mathbf{E}\|_1 - \langle \mathbf{Z}_0, \mathbf{E}\rangle \geq \mu\|\mathbf{E} - \mathbf{E}_0\|_F^2. \tag{26}$$

Thus, overall we have

$$\theta_3(x_3) \geq \theta_3(x_3^*) + \langle A_3^T\lambda^*, x_3 - x_3^*\rangle + \mu\|\mathbf{E} - \mathbf{E}_0\|_F^2. \tag{27}$$

Combining with (24), we arrive at

$$\langle y_3 - A_3^T\lambda^*, x_3 - x_3^*\rangle \geq \mu|x_3 - x_3^*|^2, \ \forall x_3 \text{ and } \forall y_3 \in \partial\theta_3(x_3), \tag{28}$$

which shows that $\|E\|_1$ satisfies the sub-strong monotonicity assumption.

Additionally, $\|\mathbf{H}\|_*, \|\mathbf{B}\|_*$ are close and proper convex and A's have full column rank. We thus deduce that the direct extension of ADMM, Algorithm 1, applied to objective (17) is convergent according to [38]. □

4 Experimental Results

4.1 Parameter calibration

In this section, we illustrate the enhancement made by side information through both numerical simulations and real-world applications. First, we explain how parameters used in our implementation are tuned. Second, we compare the recoverability of our proposed algorithms with state-of-the-art methods for incorporating features or dictionary, viz. PCPF [17] and LRR [11] on synthetic data as well as the baseline PCP [9] when there are no features available. Last, we show how powerful side information can be for the task of UV completion in post-invariant face recognition, where both features and side information are derived from generative adversarial networks.

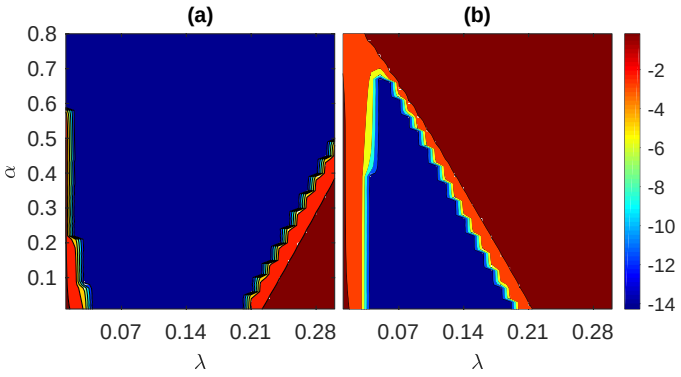


Fig. 3. Log-scale relative error ($\log \frac{\|\mathbf{L}-\mathbf{L}_0\|_F}{\|\mathbf{L}_0\|_F}$) of PCPSM **(a)** when side information is perfect ($\mathbf{S} = \mathbf{L}_0$) and **(b)** when side information is the observation ($\mathbf{S} = \mathbf{M}$).

For LRR, clean subspace X is used as in [24] instead of the observation $\mathbf{X}$ itself as the dictionary. PCP is solved via the inexact ALM [37] and the heuristics for predicting the dimension of principal singular space is not adopted here due to its lack of validity on uncharted real data [39]. We also include Partial Sum of Singular Values (PSSV) [40] in our comparison for its stated advantage in view of the limited number of images available. The stopping criteria for PCPF, LRR, PCP and PSSV are all set to the same KKT optimality conditions for reasons of consistency.

In order to tune the algorithmic parameters, we first conduct a benchmark experiment as follows: a low-rank matrix $\mathbf{L}_0$ is generated from $\mathbf{L}_0 = \mathbf{J}\mathbf{K}^T$, where $\mathbf{J}, \mathbf{K} \in \mathbb{R}^{200\times 10}$ have entries from a $\mathcal{N}(0, 0.005)$ distribution; a 200×200 sparse matrix $\mathbf{E}_0$ is generated by randomly setting 38000 entries to zero with others taking values of ± 1 with equal probability; side information $\mathbf{S}$ is assumed perfect, that is, $\mathbf{S} = \mathbf{L}_0$; $\mathbf{U}$ is set as the left-singular vectors of $\mathbf{L}_0$; and $\mathbf{V}$ is set as the right-singular vectors of $\mathbf{L}_0$; all entries are observed. It has been found that a scaling ratio $\beta = 1.1$, a tolerance threshold $\epsilon = 10^{-7}$ and a maximum step size $\mu = 10^7$ to avoid ill-conditioning can bring all models except PSSV to convergence with a recovered $\mathbf{L}$ of rank 10, a recovered $\mathbf{E}$ of sparsity 5% and an accuracy $\|\mathbf{L} - \mathbf{L}_0\|_F/\|\mathbf{L}_0\|_F$ on the order of 10^{-6}. Still, these apply to PSSV as is done similarly in [40].

Although theoretical determination of α and λ is beyond the scope of this paper, we nevertheless provide empirical guidance based on extensive experiments. A parameter weep in the $\alpha - \lambda$ space for perfect side information is shown in Figure3(a) and for observation as side information in Figure3(b) to impart a lower bound and a upper bound respectively. It can be easily seen that $\lambda = 1/\sqrt{200}$ (or $\lambda = 1/\sqrt{\max(n_1, n_2)}$ for a general matrix of dimension $n_1 \times n_2$) from Robust PCA works well in both cases. Conversely, α depends on the quality of the side information. When the side information is accurate, a large α should be selected to capitalise upon the side information as much as possible, whereas when the side information is improper, a small α should be picked to sidestep the dissonance caused by the side information. Here, we have discovered that a value of 0.2 works best with synthetic data and a value of 0.5 is suited for public video sequences, both of which will be used in all experiments in subsequent sections together with other aforementioned parameter settings. It is worth emphasising again that prior knowledge of the structural information about the data yields more appropriate values for α and λ.

4.2 Phase transition on synthetic datasets

We now focus on the recoverability problem, i.e. recovering matrices of varying ranks from errors of varying sparsity. True low-rank matrices are created via $\mathbf{L}_0 = \mathbf{J}\mathbf{K}^T$, where $200 \times r$ matrices $\mathbf{J}, \mathbf{K}$ have independent elements drawn

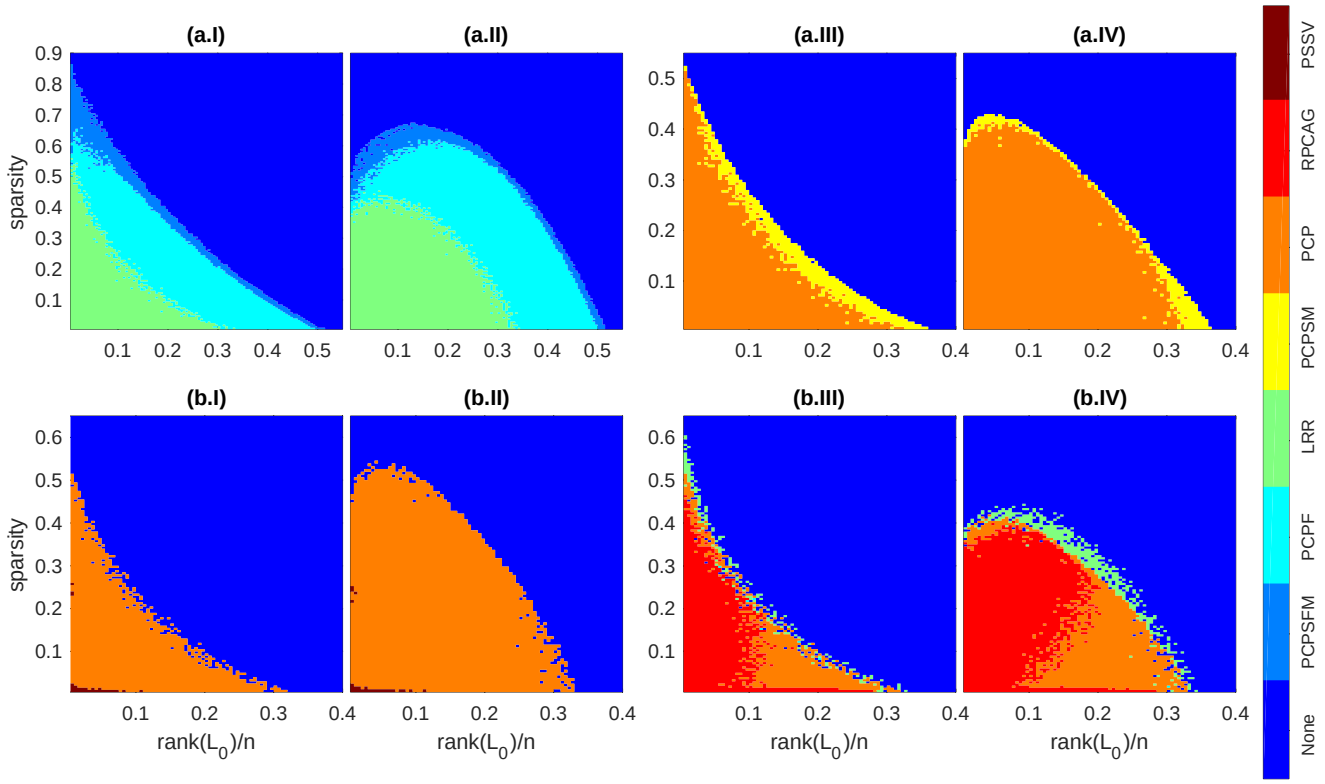


Fig. 4. Domains of recovery by various algorithms in the fully observed case: **(I,III)** for random signs and **(II,IV)** for coherent signs.

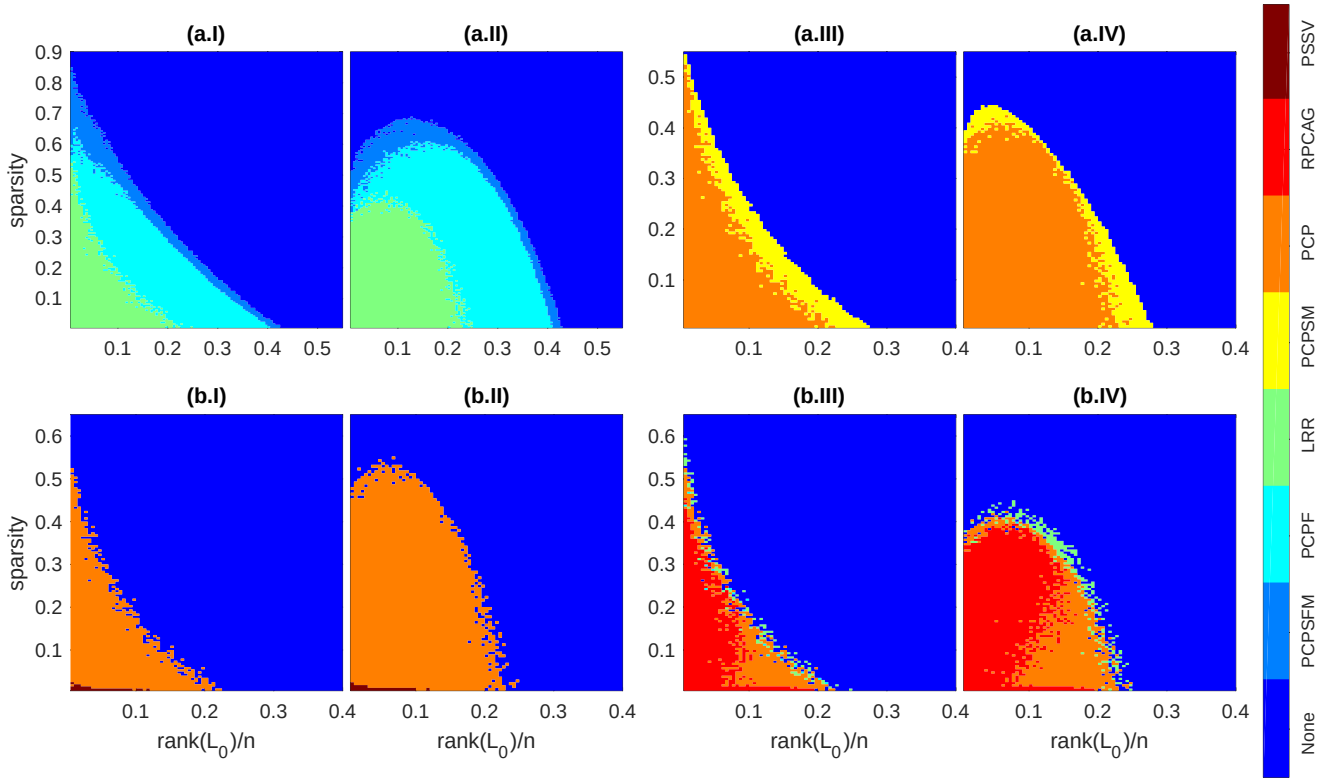


Fig. 5. Domains of recovery by various algorithms in the partially observed case: **(I,III)** for random signs and **(II,IV)** for coherent signs.

randomly from a Gaussian distribution of mean 0 and variance $5 \cdot 10^{-3}$, thus r is the rank of $\mathbf{L}_0$. Next, we generate 200×200 error matrices $\mathbf{E}_0$, which possess $\rho_s \cdot 200^2$ non-zero elements located randomly within the matrix. We consider two types of entries for $\mathbf{E}_0$: Bernoulli ± 1 and $\mathcal{P}_\Omega(\text{sgn}(\mathbf{L}_0))$, where $\mathcal{P}$ is the projection operator. $\mathbf{X} = \mathbf{L}_0 + \mathbf{E}_0$ thus becomes the simulated observation. For each (r, ρ_s) pair, three observations are constructed. The recovery is successful if for all these three problems, the following criteria regarding the recovered $\mathbf{L}$ is met:

$$\frac{\|\mathbf{L} - \mathbf{L}_0\|_F}{\|\mathbf{L}_0\|_F} < 10^{-3}. \tag{29}$$

In addition, let $\mathbf{L}_0 = \mathbf{M}\mathbf{\Sigma}\mathbf{Y}^T$ be the SVD of $\mathbf{L}_0$. Feature $\mathbf{U}$ is formed by randomly interweaving column vectors of $\mathbf{M}$ with d arbitrary orthonormal bases for the null space of $\mathbf{M}^T$, while permuting the expanded columns of $\mathbf{Y}$ with d random orthonormal bases for the kernel of $\mathbf{Y}^T$ forms feature $\mathbf{V}$. Hence, the feasibility conditions are fulfilled: $\mathbb{C}(\mathbf{U}) \supseteq \mathbb{C}(\mathbf{L}_0)$, $\mathbb{C}(\mathbf{V}) \supseteq \mathbb{C}(\mathbf{L}_0^T)$, where $\mathbb{C}$ is the column space operator.

For each trial, we construct the side information by directly adding small Gaussian noise to each element of $\mathbf{L}_0$: $l_{ij} \rightarrow l_{ij} + \mathcal{N}(0, 2.5r \cdot 10^{-9})$, $i, j = 1, 2, \cdots, 200$. As a result, the standard deviation of the error in each element is 1% of that among the elements themselves. On average, the Frobenius percent error, $\|\mathbf{S} - \mathbf{L}_0\|_F / \|\mathbf{L}_0\|_F$, is 1%. Such side information is genuine in regard to the fact that classical PCA with accurate rank is not able to eliminate the noise [41]. We set d to 10 throughout.

Full observation Figures 4 (a.I) and (a.II) plot results from PCPF, LRR and PCPSFM. On the other hand, the situation with no available features is investigated in Figures 4 (a.III) and 4 (a.IV) for PCP and PCPSM. The frontier of PCPF has been advanced by PCPSFM everywhere for both sign types. Especially at low ranks, errors with much higher density can be removed. Without features, PCPSM surpasses PCP by and large, with significantly more recovery at small sparsity levels for both sign cases. Results from RPCAG and PSSV are worse than PCP with LRR marginally improving (see Figures 4(b.I), (b.II), (b.III) and b(IV)).

Partial observation Figures 5 (a.I) and (a.II) map out the results for PCPF, LRR and PCPSFM when 10% of the elements are occluded and Figures 5 (a.III) and (a.IV) for featureless PCP and PCPSM. In all cases, areas of recovery are reduced. However, there are now larger gaps between PCPF and PCPSFM, so as for PCP and PCPSM. This marks the usefulness of side information particularly in the event of missing observations. We realise that in unrecoverable areas, PCPSM and PCPSFM still obtain much smaller values of $\|L - L_0\|_F$. FRPCAG fails to recover anything at all.

4.3 Face denoising

If a surface is convex Lambertian and the lighting is isotropic and distant, then the rendered model spans a 9-D linear subspace [42]. Nonetheless, facial images are only approximately so because facial harmonic planes have negative pixels and real lighting conditions entail unavoidable occlusion and albedo variations. It is thus more reasonable to decompose facial image formation as a low-rank component for face description and a sparse component for defects. In pursuit of this low-rank portrayal, we suggest that there can be further boost to the performance of facial characterisation by leveraging an image which faithfully represents the subject.

We consider images of a fixed pose under different illuminations from the extended Yale B database for testing. All 64 images were studied for each person. 32556×64 observation matrices were formed by vectorising each 168×192 image and the side information was chosen to be the average of all images, tiled to the same size as the observation matrix

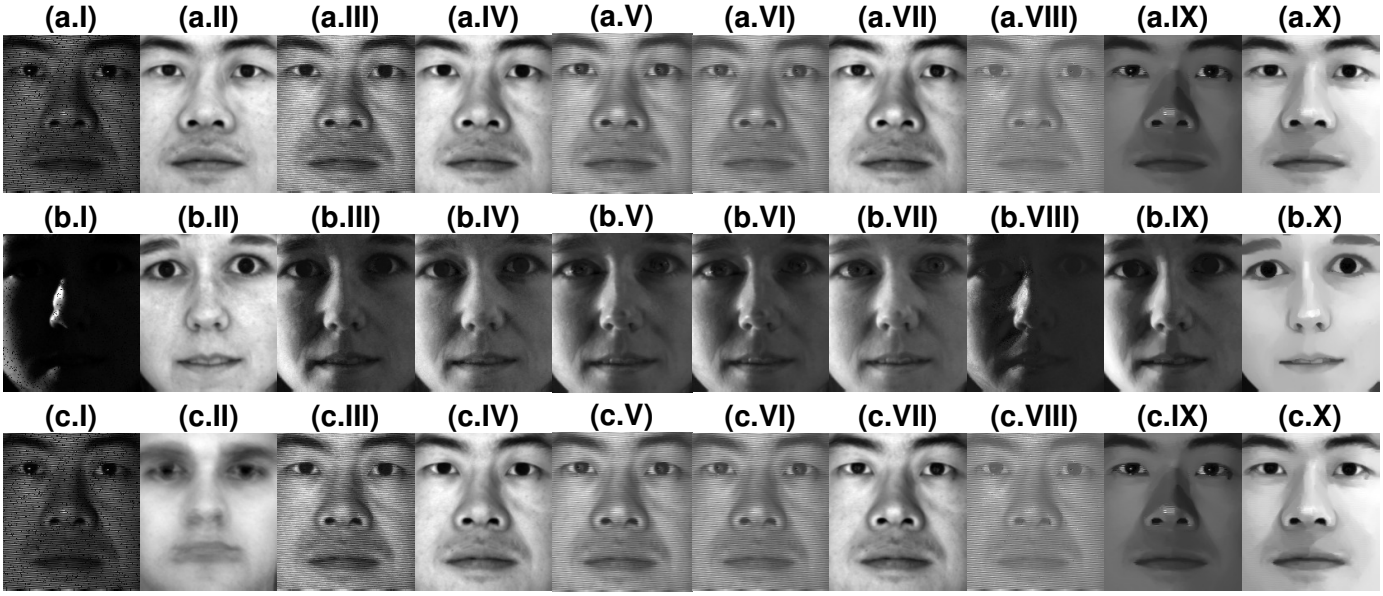


Fig. 6. Comparison of face denoising ability: **(I)** Observation; **(II)** side information; **(III)** PCP; **(IV)** PCPSM; **(V)** LRR; **(VI)** PCPF; **(VII)** PCPFSM; **(VIII)** PSSV; **(IX)** RPCAG; and **(X)** FRPCAG.

for each subject. In addition, 5% of the randomly selected pixels within each image were set as missing entries.

For LLR, PCPF and PCPSFM to run, we learn the feature dictionary following an approach by Vishal et al. [43], which is a popular method for extracting high-level attributes [44]. In a nutshell, the feature learning process can be treated as a sparse encoding problem. More specifically, we simultaneously seek a dictionary $\mathbf{D} \in \mathbb{R}^{n_1 \times c}$ and a sparse representation $\mathbf{B} \in \mathbb{R}^{c \times n_2}$ such that:

$$\min_{\mathbf{D},\mathbf{B}} \|\mathbf{M} - \mathbf{D}\mathbf{B}\|_F^2 \quad \text{s.t.} \quad \gamma_i \leq t \text{ for } i = 1 \dots n_2, \tag{30}$$

where c is the number of atoms, γ_i counts the number of non-zero elements in each sparsity code and t is the sparsity constraint factor. This can be solved by the K-SVD algorithm [45]. Here, feature $\mathbf{U}$ is the dictionary $\mathbf{D}$ and feature $\mathbf{V}$ corresponds to a similar solution using the transpose of the observation matrix as input. For implementation details, we set c to 40, t to 40 and used 10 iterations for each subject.

As a visual illustration, two challenging cases are exhibited in Figure6. For subject $\#2$, it is clearly evident that PCPSM and PCPSFM outperform the best existing methods through the complete elimination of acquisition faults. More surprisingly, PCPSFM even manages to restore the flash in the pupils that is barely present in the side information. For subject $\#34$, PCPSM indubitably reconstructs a more vivid right eye than that from PCP which is only discernible. With that being said, PCPSFM still prevails by uncovering more shadows, especially around the medial canthus of the right eye, and revealing a more distinct crease in the upper eyelid as well a more translucent iris. We further unmask the strength of PCPSM and PCPSFM by considering the stringent side information made of the average of 10 other subjects. Surprisingly, PCPSM and PCPSFM still manage to remove the noise and recover an authentic image (Figure6 (c.IV) and 6 (c.VII)). We also notice that PSSV, RPCAG, FRPCAG do not improve upon PCP as in simluation experiments. Thence, we will focus on comparisons with PCP, LRR, PCPF only.

4.4 UV map completion

We concern ourselves with the problem of completing the UV texture for each of a sequence of video frames. That is, we apply PCPSM and PCPSFM to a collection of incomplete textures lifted from a video. This parameter-free approach is advantageous to a statistical texture model such as the 3D Morphable Model (3DMM) [46], [47] by virtue of its difficulty in reconstructing unseen images captured 'in-the-wild' (using any commercial cameras in arbitrary conditions).

4.4.1 Texture extraction

Given a 2D image, we extract its UV texture by fitting the 3DMM. More specifically, following [48], three parametric models are employed. These are a 3D shape model (31), a texture model (32) and a camera model (33):

$$\mathcal{S}(\mathbf{p}) = \bar{\mathbf{s}} + \mathbf{U}_s\mathbf{p}, \tag{31}$$

$$\mathcal{T}(\boldsymbol{\lambda}) = \bar{\mathbf{t}} + \mathbf{U}_t\boldsymbol{\lambda}, \tag{32}$$

$$\mathcal{W}(\mathbf{p}, \mathbf{c}) = \mathcal{P}(\mathcal{S}(\mathbf{p}), \mathbf{c}), \tag{33}$$

where $\mathbf{p} \in \mathbb{R}^{n_s}$, $\boldsymbol{\lambda} \in \mathbb{R}^{n_t}$ and $\mathbf{c} \in \mathbb{R}^{n_c}$ are shape, texture and camera parameters to optimise; $\mathbf{U}_s \in \mathbb{R}^{3N \times n_s}$ and $\mathbf{U}_t \in \mathbb{R}^{3N \times n_t}$ are the shape and texture eigenbases respectively, with N being the number of vertices in the shape model; $\bar{\mathbf{s}} \in \mathbb{R}^{3N}$ and $\bar{\mathbf{t}} \in \mathbb{R}^{3N}$ are the corresponding means of shape and texture models, which are learnt from facial scans of 10000 individuals [47]; $\mathcal{P}(\mathbf{s}, \mathbf{c}) : \mathbb{R}^{3N} \to \mathbb{R}^{2N}$ is a perspective camera transformation function. The complete cost function for 3DMM fitting is:

$$\min_{\mathbf{p},\boldsymbol{\lambda},\mathbf{c}} \|\mathbf{F}(\mathcal{W}(\mathbf{p}, \mathbf{c})) - \mathcal{T}(\boldsymbol{\lambda})\|^2 + \beta_l \|\mathcal{W}(\mathbf{p}, \mathbf{c})) - \mathbf{s}_l\|^2 + \beta_s \|\mathbf{p}\|^2_{\boldsymbol{\Sigma}_s^{-1}} + \beta_t \|\boldsymbol{\lambda}\|^2_{\boldsymbol{\Sigma}_t^{-1}}, \tag{34}$$

where $\mathbf{F}(\mathcal{W}(\mathbf{p}, \mathbf{c}))$ denotes the operation of sampling the feature image onto the projected 2D locations. The second term is a landmark term with weighting β_l in order to accelerate in-the-wild 3DMM fitting, where the 2D shape, $\mathbf{s}_l$, is provided by [49]. The final two terms are regularisation terms to counter over-fitting, where $\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_t$ are diagonal matrices with the main diagonal being eigenvalues of the shape and texture models respectively. Eq. 34 is solved by the Gauss-Newton optimisation framework (see [48] for details). We empirically set $\beta_l = 10^5$, $\beta_s = 3 \times 10^6$ and $\beta_t = 1$ following [50], [51]. Note that any landmark localisation techniques [52] can be applied within our framework and

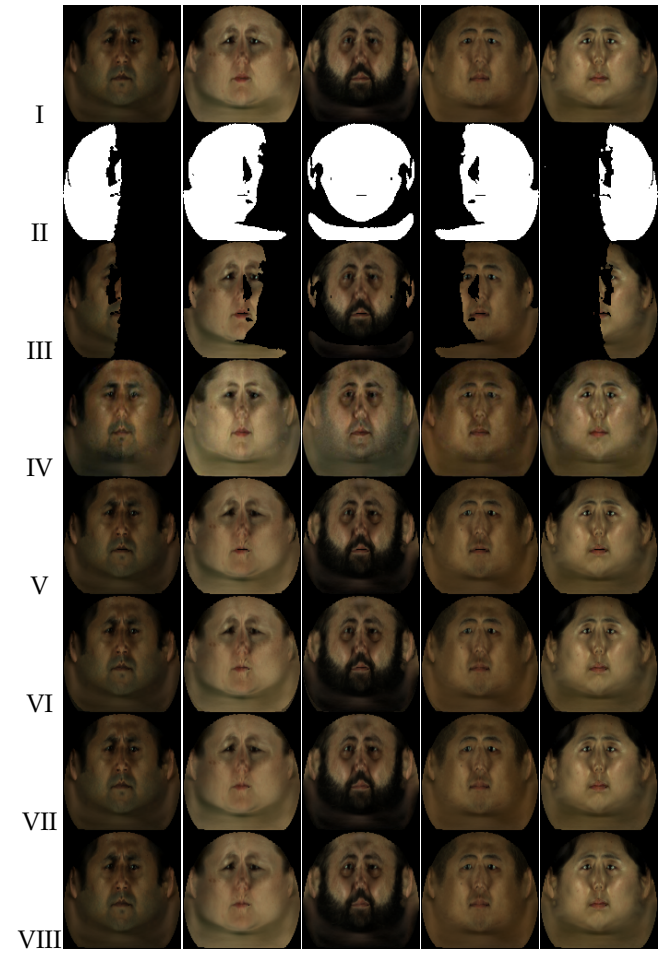


Fig. 7. (row I) original sequences; (row II) random masks; (row III) sample inputs; (row IV) side information; (row V) PCP; (row VI) PCPSM; (row VII) LRR; (row VIII) PCPSFM.

TABLE 1
Quantitative measures of UV completion by various algorithms on the 4DFAB dataset.

Subject		#1	#2	#3	#4	#5
PSNR (dB)	PCP	35.99 ±0.79	26.75 ±0.88	32.65 ±0.88	31.33 ±0.99	29.10 ±1.68
	PCPSM	**39.56** ±1.30	**30.63** ±1.47	**34.66** ±1.29	**35.86** ±1.85	**32.80** ±2.93
	LRR	40.94 ±2.13	30.69 ±1.71	36.38 ±2.10	35.94 ±2.53	33.97 ±3.93
	PCPSFM	**41.48** ±2.06	**31.46** ±1.69	**37.29** ±2.37	**36.60** ±2.36	**34.80** ±4.14
SSIM	PCP	0.973 ±0.004	0.922 ±0.012	0.962 ±0.010	0.956 ±0.007	0.949 ±0.013
	PCPSM	**0.987** ±0.004	**0.952** ±0.013	**0.969** ±0.010	**0.981** ±0.006	**0.973** ±0.013
	LRR	0.990 ±0.005	0.952 ±0.013	0.975 ±0.010	0.982 ±0.007	0.978 ±0.014
	PCPSFM	**0.991** ±0.004	**0.958** ±0.013	**0.979** ±0.010	**0.984** ±0.007	**0.981** ±0.013

the visible mask of facial region is a natural product of the 3DMM fitting process.

4.4.2 Quantitative evaluation

We quantitatively evaluate the completed UV maps by our proposed methods on the 4DFAB dataset [53]. 4DFAB is the first 3D dynamic facial expression dataset designed for biometric applications, where 180 participants are invited to attend four sessions at different times. Hence, to complete UV maps for one session, we can leverage images from another session as side information. For each of 5 randomly selected subjects, one dynamic sequence of 155 frames is randomly cut from the second session. After vectorisation, a 32556×155 observation matrix is formed. To produce UV masks of different poses, we rotate each face with different yaw and pitch angles. The yaw angle ranges from -90° to 90° in steps of 6°, whereas the pitch angle is selected from $\{-10^\circ, -5^\circ, 0^\circ, 5^\circ, 10^\circ\}$. Therefore, for each subject, a set of 155 unique masks are generated. We also tiled one image of the same subject from the first session into a 32556×155 matrix as side information. $\mathbf{U}$ is provided by the left singular vector of the original sequence while $\mathbf{V}$ is set to the identity.

From Figure 7, we observe that (I) RPCA approaches can deal with cases where more than 50% of the pixels are missing; (II) imperfect side information (shaved beard, removed earrings and different lightings) still help with the recovery process. We record peak signal-to-noise ratios (PSNR) and structural similarity indices (SSIM) between the completed UV maps and the original maps in Table 1. It is evident that with the assistance of side information, much higher fidelity can be achieved. The use of imperfect side information nearly comes on a par with perfect features.

4.4.3 Generative adversarial networks

More often than not, ground-truth $\mathbf{U}$, $\mathbf{V}$ are not accessible to us for in-the-wild videos. Learning methods such as (30) must be leveraged to acquire $\mathbf{U}$ or $\mathbf{V}$. However, (30) is not ideal: (I) it is not robust to errors of arbitrary magnitude; (II) it cannot handle missing values; (III) it requires exhaustive search of optimal parameters which vary from video to video; (IV) it only admits greedy solutions[1]. As a matter of fact, we can use GAN to produce authentic pseudo ground-truth. Then we apply truncated singular value decomposition to the vectorised frames and use the obtained left and right singular vectors as $\mathbf{U}$ and $\mathbf{V}$ subspace features respectively. Moreover, such completed sequence provides us with good side information. For each color channel, we average the video frames before tiling it back to the original length. This resulting matrix is taken as side information. For GAN, we employ the image-to-image conditional adversarial network [55] (appropriately customised) to conduct UV completion. Details regarding the architecture and training of GAN can be found in the supplementary materials.

4.4.4 Qualitative demonstration

To examine the ability of our proposed methods on in-the-wild images. We perform experiments on the 300VW dataset [56]. This dataset contains 114 in-the-wild videos that exhibit large variations in pose, expression, illumination, background, occlusion, and image quality. Each video shows exactly one person, and each frame is annotated with 68 facial landmarks. We perform 3DMM fitting on these videos and lift one corresponding UV map for each frame, where the visibility mask is produced by z-buffering based on the fitted mesh. Side information is generated by taking the average of the completed UVs from GAN. $\mathbf{U}$ and $\mathbf{V}$ are assigned to the singular vectors of the completed texture sequence from GAN.

We display results for one sample frame from each of 9 arbitrary videos in Figure 9 of the supplementary materials. As evident from the images, GAN alone has unavoidable drawbacks: (I) when 3DMM fitting is not accurate, GAN is unable to correct such defects; (II) when the image itself contains errors, GAN is unable to remove them. On the other hand, PCP often fails to produce a complete UV. PCPSM always produces a completed UV texture, which is an improvement over PCP, but it generates undesirable boundaries. Visually, LRR and PCPSFM have the best performance, being able to produce good completed UVs for a large variety of poses, identities, lighting conditions and facial characteristics. This justifies the quality of subspaces and side information from GAN for use in the robust PCA framework. We also synthesise 2D faces of three different poses using the the completed UV maps in Figure8.

4.5 Face recognition

Face recognition is a crucial element of biometrics [57], [58], [59], [60], [61], [62], [63]. In this paper, we focus on the set-based face verification, i.e. to decide whether two sets of facial images are of the same person or not. One face set could consist of one or multiple samples of the same person (e.g. still images, or frames from a video of the person, or a mixture of both). Therefore, traditional face verification is a special case of the set-based face verification.

The simplest approach to the set-based face verification problem is to generate a feature vector per image, aggregate

1. There is a variant of KSVD [54] that can fill holes which are smaller than the size of the atoms. We evaluate it against our GAN-based approach in Figure 3 and Table 1 of the supplementary materials.

Fig. 8. 2D face synthesis of three views ($-45°$, $0°$, $45°$) from the completed UV maps by various methods.

them into one vector to represent the set (e.g. calculate the feature centre by average), and then compute the cosine similarity between sets. However, the combination rule of averaging is oversimplified since not all face images in one set are of equal importance. The features derived from a profile face is probably of less importance than the features coming from a frontal face as there is signal loss due to self-occlusion under pose variations.

More specifically, we focus on pose-invariant face recognition. Modern approaches to pose-invariant face recognition include pose-robust feature extraction [64], multi-view subspace learning [65], face frontalisation by synthesis [51], etc [66]. Nonetheless, these methods often fall short of expectations either due to fundamental limitations or inability to fuse with other useful methods. For example, Generalised Multi-view Analysis [67] cannot take account of pose normalisation [68] or deep neural network-based pose-robust feature extraction [69], and vice versa. Hence, it is fruitful to provide a framework where information from different perspectives can be fused together to deliver better prediction.

We quantitatively evaluate our proposed fusion methods by carrying out set-based face verification experiments. The experiments are performed on four standard databases, namely CFP [70], IJB [71], [72], [73], YTF [74] and PaSC [75]. Evaluation results on these benchmarks will be given in the next few sections. Overall, the proposed method outperforms current state-of-the-art approaches [59], [60], [76], [77], [78] by a large margin.

4.5.1 Face Feature Embedding

We employ ArcFace [79] with ResNet50 [80] as the backbone. The additive angular margin loss ($m = 0.35$) is used to train a 512-D facial feature embedding model on the VGG2 training set [76], which contains 3,141,890 images from 8,631 identities. Following [79], we use five facial landmarks (eye centres, nose tip and mouth corners) [81] to normalise the face images by similarity transformation. The faces are cropped and resized to 112×112. Figure 9 illustrates the set-based face feature embedding used for face verification. For one facial image set, we first extract 3D face shapes and incomplete UV maps via 3DMM fitting [48]. Then, we utilise the proposed UV completion methods (GAN [51], PCP, PCPSM, LRR and PCPSFM) to derive completed UV maps. Frontal faces are synthesised from the full UV maps and the 3D shapes, which are then fed into the feature embedding network. A set of 512-D features from the last fully connected layer of network, is used to compute the feature centre and eventually taken as the feature descriptor.

4.5.2 Evaluation Metrics

In this paper, we employ the standard 1:1 verification protocol. The performance is reported by the true accept (positive) rates (TAR) vs. false accept (positive) rates (FAR)

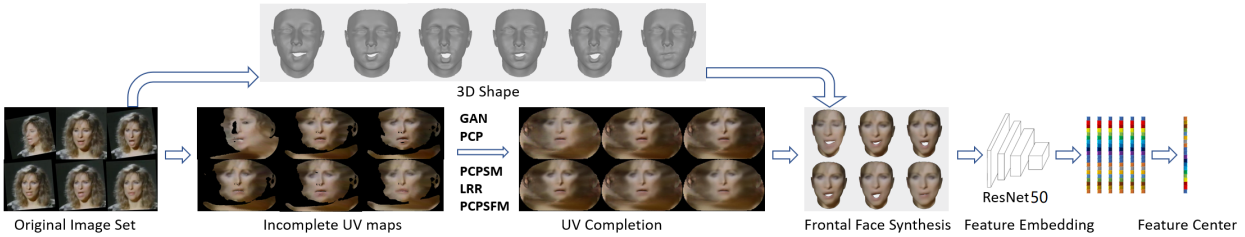


Fig. 9. The proposed pipeline for video-based face recognition. The 3DMM [48] is fitted on the frames of the video and the incompleted UV maps are estimated. The trained GAN [51] is then used to provide an initial estimate of the side information and the proposed methodology is applied to generate the completed UV maps. The 3D model is reused to render the images in the frontal view. Deep neural network is used to extract features from all frames and the average of the features is used to represent the video.

(from the receiver operating characteristics (ROC) curve). Following [78], we are interested in the TAR values where FAR=1e-4 and FAR=1e-5, which is also the security level for financial applications. Apart from the ROC curve, we also calculate the best threshold value from the positive and negative pairs, and report the corresponding classification accuracy for each method on the YTF dataset.

4.5.3 Ablation Experiments on CFP

The CFP dataset [70] consists of 500 subjects, each of which has 10 frontal and 4 profile images. For each subject, we construct four sets (with 3, 3, 4 and 4 faces respectively) where each set includes at least one profile face. For set-based face verification on CFP, we extensively compare all possible $3,000$ positive pairs and $1,996$K negative pairs.

As shown in Table 2 and Figure 10, we compare the proposed methods with several baseline methods. It can be clearly observed that by leveraging subspace features or side information from GAN (LRR/PCPSM), we ameliorate the recognition results in terms of TAR over the vanilla PCP, while a further boost in performance can be achieved when both of them are considered together (PCPSFM). Compared to the result of ArcFace, the proposed PCPSFM achieves a TAR improvement of 1.7% at FAR=1e-5.

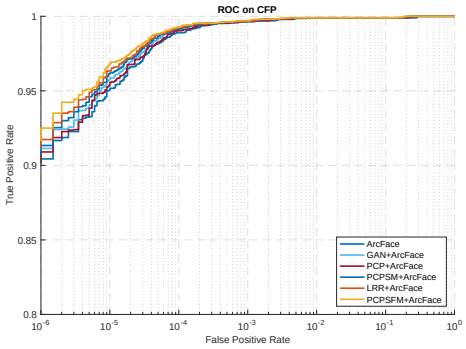


Fig. 10. ROC curves on the CFP dataset.

4.5.4 Experiments on IJB

The IARPA Janus Benchmarks have been gradually enlarged from IJB-A [71] to IJB-B [72] and IJB-C [73]. The IJB-A dataset contains $5,712$ images and $2,085$ videos from 500 subjects, with an average of 11.4 images and 4.2 videos per subject.

TABLE 2
Verification TAR on the CFP dataset, the higher TAR the better.

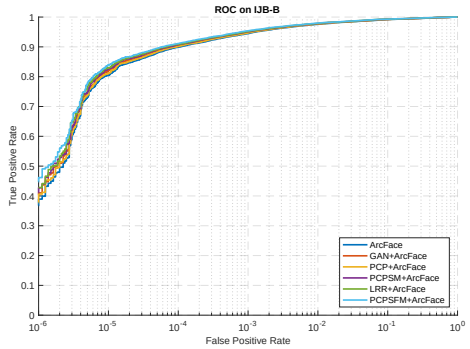
Method	FAR=1e-6	FAR=1e-5	FAR=1e-4
ArcFace	0.901	0.950	0.989
GAN+ArcFace	0.905	0.957	0.991
PCP+ArcFace	0.902	0.953	0.990
LRR+ArcFace	0.911	0.963	0.993
PCPSM+ArcFace	0.907	0.961	0.991
PCPSFM+ArcFace	0.916	0.967	0.993

The IJB-B dataset is an extension of IJB-A, which contains $1,845$ subjects with 21.8K still images and 55K frames from $7,011$ videos. In total, there are $12,115$ templates with $10,270$ genuine matches and 8M impostor matches. The IJB-C dataset is a further extension of IJB-B, having $3,531$ subjects with 31.3K still images and 117.5K frames from $11,779$ videos. In total, there are $23,124$ templates with $19,557$ genuine matches and $15,639$K impostor matches. All images and videos from the IARPA Janus Benchmarks are captured under unconstrained environment and show large variations in expression and image qualities. Since IJB-A has been superseded by IJB-B with its images being a subset of IJB-B, we only report the results on IJB-B and IJB-C.

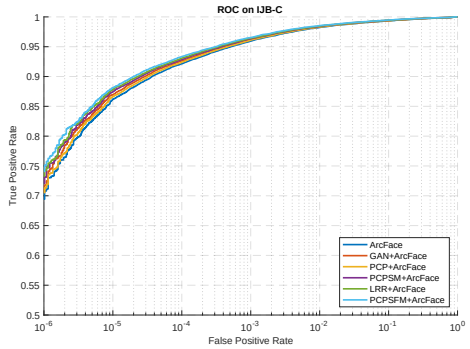
In Figure11, we illustrate the ROC curves of the proposed method against the baselines. We see that ArcFace [79] achieves strong performance. However, PCPSFM further increases the performance through incorporating feature subspace and side information even when there are some low-resolution face images within the template. This is because the proposed method can integrate information from different face images within the template and therefore make the final template feature representation robust. To conduct fair comparison with other methods [76], [77], [78], no flip test and face detection scores are used during evaluation even though both tricks are known to improve the performance.

In Table 3 and 4, comparisons between the proposed PCPSFM and the most recent methods [76], [77], [78], [82], [83] are made. We can see from the results that the baseline method, ArcFace [79], already achieves similar or even better performance compared to the methods proposed in [82], [83]. With the assistance of the proposed PCPSFM, our method achieves the best result on both IJB-B and IJB-C datasets outperforming counterparts [82], [83] even with less identities in the training data and a smaller CNN

embedding network.



(a) ROC for IJB-B



(b) ROC for IJB-C

Fig. 11. ROC curves of 1:1 verification protocol on the IJB-B and IJB-C dataset.

TABLE 3
1:1 verification TAR on the IJB-B dataset (Higher is better).

Method	FAR=1e-4	FAR=1e-3
GOTS [72]	0.160	0.330
VGGFaces [60], [72]	0.550	0.720
FPN [83]	0.832	0.916
Light CNN [84]	0.877	0.920
Centre Loss [85]	0.807	0.900
Crystal Loss [82]	0.898	0.944
Whitelam et al. [72]	0.540	0.700
Navaneeth et al. [86]	0.685	0.830
ResNet50 [76]	0.784	0.878
SENet50 [76]	0.800	0.888
ResNet50+SENet50 [76]	0.800	0.887
MN-v [77]	0.818	0.902
MN-vc [77]	0.831	0.909
ResNet50+DCN(Kpts) [78]	0.850	0.927
ResNet50+DCN(Divs) [78]	0.841	0.930
SENet50+DCN(Kpts) [78]	0.846	0.935
SENet50+DCN(Divs) [78]	0.849	0.937
ArcFace [79]	0.899	0.945
GAN+ArcFace	0.904	0.949
PCP+ArcFace	0.901	0.947
PCPSM+ArcFace	0.907	0.951
LRR+ArcFace	0.909	0.952
PCPSFM+ArcFace	**0.911**	**0.954**

4.5.5 Experiments on YTF

The YouTube Face (YTF) dataset [74] consists of $3,425$ videos from $1,595$ different people. The clip duration varies from 48 frames to $6,070$ frames. The average length is 181.3 frames. We follow the *unrestricted with labelled outside data* protocol and report the results on $5,000$ video pairs ($2,500$ positive pairs and $2,500$ negative pairs).

TABLE 4
1:1 verification TAR on the IJB-C dataset (Higher is better).

Method	FAR=1e-4	FAR=1e-3
Centre Loss [85]	0.853	0.912
Crystal Loss [82]	0.919	0.957
GOTS [72]	0.160	0.320
FaceNet [59]	0.490	0.660
VGG [60]	0.600	0.750
ResNet50 [76]	0.825	0.900
SENet50 [76]	0.840	0.910
ResNet50+SENet50 [76]	0.841	0.909
MN-v [77]	0.852	0.920
MN-vc [77]	0.862	0.927
ResNet50+DCN(Kpts) [78]	0.867	0.940
ResNet50+DCN(Divs) [78]	0.880	0.944
SENet50+DCN(Kpts) [78]	0.874	0.944
SENet50+DCN(Divs) [78]	0.885	0.947
ArcFace [79]	0.921	0.959
GAN+ArcFace	0.926	0.962
PCP+ArcFace	0.924	0.961
PCPSM+ArcFace	0.928	0.963
LRR+ArcFace	0.931	0.964
PCPSFM+ArcFace	**0.934**	**0.965**

This dataset is very challenging not only due to the rich pose variations but also the serious compression artifacts. We compare the performance of the proposed method with current state-of-the-art approaches on the YTF dataset. In Table 5, we list the verification accuracy for the best-performing deep learning methods. We see that our GAN model alone is among the best reported architectures and it outperforms the classical PCP. Nonetheless, their fusion (PCPSM, LRR and PCPSFM) is superior to either of them. More specifically, PCPSM improves PCP and GAN by 0.12% and 0.06% respectively. Regarding LRR, the improvements over PCP and GAN are 0.16% and 0.10% respectively. Overall, PCPSFM achieves the best result, i.e., 0.12% over PCPSM and 0.08% over LRR. We also plot the ROC curves for these methods in Figure 12. In Table 6, we list the TAR values under different FAR values. The proposed PCPSFM achieves highest TAR (83.0%) at FAR=1e-3. Arguably, the proposed PCPSFM does improve the accuracy of video-based face verification.

TABLE 5
Verification accuracy (%) of different methods on the YTF dataset.

Methods	Images	Acc (%)
DeepID [58]	0.2M	93.20
VGG Face [60]	2.6M	**97.30**
Deep Face [57]	4M	91.40
FaceNet [59]	200M	95.10
Center Loss [85]	0.7M	94.9
Range Loss [87]	1.5M	93.70
Sphere Loss [88]	0.5M	95.0
Marginal Loss [89]	4M	95.98
ArcFace	3.1M	97.52
GAN+ArcFace	3.1M	97.66
PCP+ArcFace	3.1M	97.60
PCPSM+ArcFace	3.1M	97.72
LRR+ArcFace	3.1M	97.76
PCPSFM+ArcFace	3.1M	**97.84**

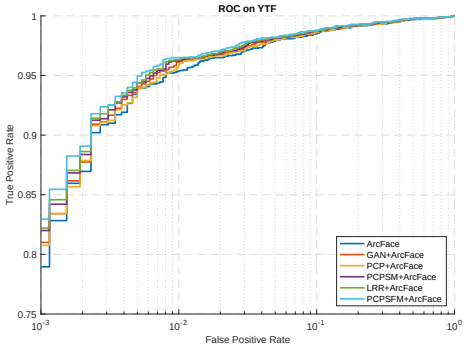


Fig. 12. ROC curves of the proposed methods on the YouTube Faces database under the "restricted" protocol.

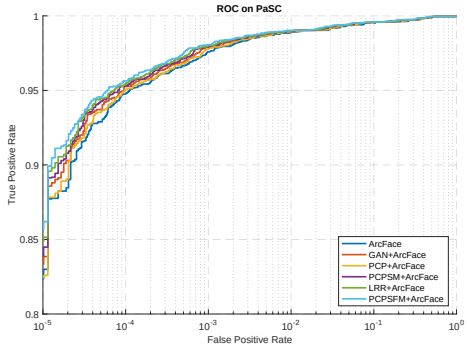


Fig. 13. ROC curves of the proposed methods on the PaSC dataset.

TABLE 6
Verification TAR on the YTF dataset (Higher is better).

Method	FAR=1e-3	FAR=1e-2	FAR=1e-1
ArcFace	0.790	0.953	0.985
GAN+ArcFace	0.810	0.961	0.987
PCP+ArcFace	0.807	0.957	0.986
PCPSM+ArcFace	0.820	0.962	0.987
LRR+ArcFace	0.822	0.963	0.987
PCPSFM+ArcFace	**0.830**	**0.965**	**0.988**

TABLE 7
Verification TAR on the PaSC dataset (Higher is better).

Method	FAR=1e-5	FAR=1e-4	FAR=1e-3
ArcFace	0.824	0.948	0.976
GAN+ArcFace	0.833	0.953	0.979
PCP+ArcFace	0.824	0.950	0.978
PCPSM+ArcFace	0.839	0.953	0.979
LRR+ArcFace	0.849	0.954	0.980
PCPSFM+ArcFace	**0.857**	**0.956**	**0.981**

4.5.6 Experiments on PaSC

The PaSC dataset [75] includes $9,376$ still images and $2,802$ videos from 293 people. The images are evenly split with respect to the distance to the camera, alternative sensors, frontal versus not-frontal views and different environments. There are three protocols for face verification: comparing still images to still images, videos to videos, and still images to videos. Since we have conducted image-to-image and video-to-video experiments in previous sections, we only report image-to-video results on PaSC with the public evaluation toolkit.

As the PaSC dataset [75] includes static images and videos of the same people, it is very interesting to explore face verification performance between modalities: static image to dynamic video. Simply put, given only a few images of a person, can we verify this person in the subsequent video that he/she is seen or claimed to be seen? To set up this experiment, we prepare a query set of $1,401$ handheld (or alternatively controlled) videos and a target set comprising of $9,376$ still images from 293 identities. Figure 13 presents the ROC curve of each method. In Table 7, we report the TAR at different FARs. The proposed PCPSFM significantly improves TAR from 82.4% to 85.7% at FAR=1e-5. In [75], the baseline method only obtains TAR of 42% at FAR=1e-2, whereas our method PCPSFM achieves TAR of 99.0% at FAR=1e-2.

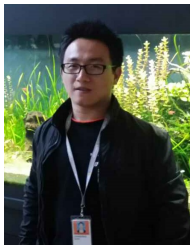
5 Conclusions

In this paper, we study the problem of robust principal component analysis with features acting as side information in the presence of missing values. For the application domain of UV completion, we also propose the use of generative adversarial networks to extract side information and subspaces, which, to the best of our knowledge, is the first occasion where RPCA and GAN have been fused. We also prove the convergence of ADMM for our convex objective. Through synthetic and real-world experiments, we demonstrate the advantages of side information. In virtue of in-the-wild data, we corroborate our fusion strategy. Finally, face recognition benchmarks accredit the efficacy of our proposed approach over state-of-the-art methods. Further works include extending our approaches to new application domains, such as pose estimation and gender estimation [90].

6 Acknowledgements

This work was partially funded by the EPSRC project EP/N007743/1 (FACER2VM: Face Matching for Automatic Identity Retrieval, Recognition, Verification and Management), the EPSRC project EP/S010203/1 (DEFORM: Large Scale Shape Analysis of Deformable Models of Humans), the European Community Horizon 2020 [H2020/2014-2020] under grant agreement no. 688520 (TeSLA), and a Google Faculty Fellowship to Dr. Zafeiriou. We thank the NVIDIA Corporation for donating several GPUs used in this work.

Niannan Xue received the BA degree (first class) in theoretical physics from Cambridge University in 2013, and the MMath degree in applied mathematics from Cambridge University in 2014. He is currently working toward the PhD degree at Imperial College London. He was a visiting student in the Biological and Soft Systems Sector of the Cavendish Laboratory. He received St Catharine's Skerne Prize for three consecutive times. His research interests include data mining, machine learning and artificial intelligence. He is a student member of the IEEE.



Jiankang Deng is a Ph.D. candidate in the Intelligent Behaviour Understanding Group (IBUG), Department of Computing, Imperial College London. He is funded by the Imperial President's PhD Scholarships and his research interest is face analysis.

Shiyang Cheng received his B.S. degree from Northeastern University, China in 2011, and M.Sc. degree in computer science from Imperial College London in 2012. He is currently completing his Ph.D. on the topic of Robust Deformable Model for 3D Face Alignment at Imperial College London. His research interest lies on computer vision and graphics.

Yannis Panagakis is a Research Fellow in the Department of Computing, Imperial College London and a Lecturer at Middlesex University. He received his PhD and MSc degrees from the Department of Informatics, Aristotle University of Thessaloniki and his B.Sc. degree in Informatics and Telecommunication from the National and Kapodistrian University of Athens, Greece. Yannis received various scholarships and awards for his studies and research, including the prestigious Marie-Curie Fellowship in 2013. His current research interests include machine learning, signal processing, and mathematical optimization with applications to computer vision, human behaviour analysis, and music information research.

Stefanos Zafeiriou 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 Presidents 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 CVPR17 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 7000 citations to his work, h-index 44. He is the General Chair of BMVC 2017.

REFERENCES

[1] J. Booth and S. Zafeiriou, "Optimal uv spaces for facial morphable model construction," in *ICIP*, 2014.

[2] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *SIGGRAPH*, 1999.

[3] A. Patel and W. A. P. Smith, "3d morphable face models revisited." in *CVPR*, 1999.

[4] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou, "3d face morphable models in-the-wild," in *CVPR*, 2017.

[5] Q. Cao, L. Shen, W. Xie, O. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose andage," in *arXiv:1710.08092*, 2017.

[6] F. Shang, Y. Liu, J. Cheng, and H. Cheng, "Robust principal component analysis with missing data," in *CIKM*, 2014, pp. 1149–1158.

[7] C.-H. Chen, V. M. Patel, and R. Chellappa, "Learning from ambiguously labeled face images," *TPAMI*, vol. 40, no. 7, pp. 1653–1667, 2018.

[8] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[9] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[10] A. Aravkin, S. Becker, V. Cevher, and P. Olsen, "A variational approach to stable principal component pursuit," in *UAI*, 2014, pp. 32–41.

[11] B. Bao, G. Liu, C. Xu, and S. Yan, "Inductive robust principal component analysis," *TIP*, vol. 21, no. 8, pp. 3794 – 3800, 2012.

[12] R. Cabral, F. De la Torre, J. Costeira, and A. Bernardino, "Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition," in *ICCV*, 2013.

[13] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," *TIT*, vol. 58, no. 5, pp. 3047–3064, 2012.

[14] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *ISIT*, 2010.

[15] G. Liu, Q. Liu, and X. Yuan, "A new theory for matrix completion," in *NIPS*, 2017.

[16] G. Liu and P. Li, "Low-rank matrix completion in the presence of high coherence," *TSP*, vol. 64, no. 21, pp. 5623–5633, 2016.

[17] J. Jiao, T. Courtade, K. Venkat, and T. Weissman, "Justification of logarithmic loss via the benefit of side information," *TIT*, vol. 61, no. 10, pp. 5357–5365, 2015.

[18] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *TIT*, vol. 22, no. 1, pp. 1–10, 1976.

[19] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[20] K. Chiang, C. Hsieh, and I. Dhillon, "Matrix completion with noisy side information," in *NIPS*, 2015.

[21] M. Xu, J. R, and Z. Zhou, "Speedup matrix completion with side information: Application to multi-label learning," in *NIPS*, 2013.

[22] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Strategies, geometry, and bounds," *TIT*, 2017.

[23] K. Chiang, C. Hsieh, and I. Dhillon, "Robust principal component analysis with side information," in *ICML*, 2016.

[24] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Raps: Robust and efficient automatic construction of person-specific deformable models," in *CVPR*, 2014.

[25] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *ICML*, 2010, pp. 663–670.

[26] G. Liu, Q. Liu, and P. Li, "Blessing of dimensionality: Recovering mixture data via dictionary pursuit," *TPAMI*, vol. 39, no. 1, pp. 47–60, 2017.

[27] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *TPAMI*, vol. 35, no. 1, pp. 171–184, 2013.

[28] G. Liu, H. Xu, J. Tang, Q. Liu, and S. Yan, "A deterministic analysis for lrr," *TPAMI*, vol. 38, no. 3, pp. 417–430, 2016.

[29] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. Efros, "Context encoders: Feature learning by inpainting," in *CVPR*, 2016.

[30] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *CVPR*, 2017.

[31] Y. Li, S. Liu, J. Yang, and M. Yang, "Generative face completion," in *CVPR*, 2017.

[32] N. Xue, Y. Panagakis, and S. Zafeiriou, "Side information in robust principal component analysis: Algorithms and applications," in *ICCV*, 2017.

[33] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis, "Low-rank matrix recovery from errors and erasures," *TIT*, vol. 59, no. 7, pp. 4324–4337, 2013.

[34] K.-C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 615-640, p. 15, 2010.

[35] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM journal on control and optimization*, vol. 14, no. 5, pp. 877–898, 1976.

[36] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[37] Z. lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *UIUC Technical Report*, 2009.

[38] H. Sun, J. Wang, and T. Deng, "On the global and linear convergence of direct extension of ADMM for 3-block separable convex minimization models," *Journal of Inequalities and Applications*, p. 227, 2016.

[39] M. Hintermüller and T. Wu, "Robust principal component pursuit via inexact alternating minimization on matrix manifolds," *Journal of Mathematical Imaging and Vision*, vol. 51, no. 3, pp. 361–377, 2015.

[40] T. Oh, Y. Tai, J. Bazin, H. Kim, and I. Kweon, "Partial sum minimization of singular values in robust PCA: Algorithm and applications," *TPAMI*, vol. 38, no. 4, pp. 744–758, 2016.

[41] A. Shabalin and A. Nobel, "Reconstruction of a low-rank matrix in the presence of gaussian noise," *Journal of Multivariate Analysis*, vol. 118, pp. 67–76, 2013.

[42] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *TPAMI*, vol. 25, no. 2, pp. 218–233, 2003.

[43] V. Patel, T. Wu, S. Biswas, P. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *TIFS*, vol. 7, no. 3, pp. 954–965, 2012.

[44] J. Zheng and Z. . R. C. Jiang, "Submodular attribute selection for visual recognition," *TPAMI*, vol. 39, no. 11, pp. 2242–2255, 2017.

[45] M. Aharon, M. Elad, and A. Bruckstein, "rmk-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *TSP*, vol. 54, no. 11, pp. 4311–4322, 2006.

[46] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *TPAMI*, vol. 25, no. 9, pp. 1063–1074, 2003.

[47] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3d morphable model learnt from 10,000 faces," in *CVPR*, 2016.

[48] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou, "3d face morphable models in-the-wild," in *CVPR*, 2017.

[49] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d and 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)," in *ICCV*, 2017.

[50] J. Booth, A. Roussos, E. Ververas, E. Antonakos, S. Poumpis, Y. Panagakis, and S. P. Zafeiriou, "3d reconstruction of" in-the-wild" faces in images and videos," *IEEE T-PAMI*, 2018.

[51] J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, "Uv-gan: Adversarial facial uv map completion for pose-invariant face recognition," in *CVPR*, 2018.

[52] A. Kumar and R. Chellappa, "Disentangling 3d pose in a dendritic cnn for unconstrained 2d face alignment," in *CVPR*, 2018.

[53] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou, "4dfab: A large scale 4d facial expression database for biometric applications," in *CVPR*, 2018.

[54] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *TIP*, vol. 17, no. 1, pp. 53–69, 2008.

[55] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *arXiv:1611.07004*, 2016.

[56] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic, "The first facial landmark tracking in-the-wild challenge: Benchmark and results," in *ICCVW*, 2015, pp. 1003–1011.

[57] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *CVPR*, 2014.

[58] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *NIPS*, 2014, pp. 1988–1996.

[59] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015.

[60] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition." in *BMVC*, vol. 1, no. 3, 2015, p. 6.

[61] J. Chen, V. Patel, L. Liu, V. Kellokumpu, G. Zhao, M. Pietikäinen, and R. Chellappa, "Robust local features for remote face recognition," *Image and Vision Computing*, 2017.

[62] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," in *arXiv:1703.09507*, 2017.

[63] S. Shekhar, V. M. Patel, and R. Chellappa, "Synthesis-based robust low resolution face recognition," in *arXiv:1707.02733*, 2017.

[64] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 3, pp. 518–531, 2016.

[65] M. Du, A. C. Sankaranarayanan, and R. Chellappa, "Robust face recognition from multi-view videos," *IEEE transactions on image processing*, vol. 23, no. 3, pp. 1105–1117, 2014.

[66] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *TIST*, vol. 7, no. 3, pp. 1–42, 2016.

[67] A. Sharma, A. Kumar, H. Daume, and D. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *CVPR*, 2012.

[68] C. Ding, J. Choi, D. Tao, and L. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *TPAMI*, 2015.

[69] M. Kan, S. Shan, H. Chang, and X. Chen, "Stacked progressive auto-encoders (spae) for face recognition," in *CVPR*, 2014.

[70] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *WACV*, 2016.

[71] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *CVPR*, 2015, pp. 1931–1939.

[72] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. C. Adams, T. Miller, N. D. Kalka, A. K. Jain, J. A. Duncan, K. Allen *et al.*, "Iarpa janus benchmark-b face dataset." in *CVPRW*, 2017, pp. 592–600.
[73] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, and J. Cheney, "Iarpa janus benchmark–c: Face dataset and protocol." in *ICB*, 2018.
[74] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *CVPR*, 2011.
[75] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, and K. W. Bowyer, "The challenge of face recognition from digital point-and-shoot cameras," in *BTAS*, 2013, pp. 1–8.
[76] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *FG*, 2018.
[77] W. Xie and A. Zisserman, "Multicolumn networks for face recognition," *BMVC*, 2018.
[78] W. Xie, S. Li, and A. Zisserman, "Comparator networks," *ECCV*, 2018.
[79] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *arXiv:1801.07698*, 2018.
[80] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.
[81] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *SPL*, vol. 23, no. 10, pp. 1499–1503, 2016.
[82] R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J.-C. Chen, C. D. Castillo, and R. Chellappa, "A fast and accurate system for face detection, identification, and verification," *arXiv:1809.07586*, 2018.
[83] F.-J. Chang, A. T. Tran, T. Hassner, I. Masi, R. Nevatia, and G. Medioni, "Faceposenet: Making a case for landmark-free face alignment," in *ICCVW*, 2017.
[84] X. Wu, R. He, Z. Sun, and T. Tan, "A light cnn for deep face representation with noisy labels," *TIFS*, 2018.
[85] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*, 2016.
[86] N. Bodla, J. Zheng, H. Xu, J.-C. Chen, C. Castillo, and R. Chellappa, "Deep heterogeneous feature fusion for template-based face recognition," in *WACV*, 2017, pp. 586–595.
[87] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tail," in *ICCV*, 2017.
[88] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, 2017.
[89] J. Deng, Y. Zhou, and S. Zafeiriou, "Marginal loss for deep face recognition," in *CVPRW*, 2017.
[90] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *TPAMI*, 2018.