Identity-Preserving Pose-Robust Face Hallucination Through Face Subspace Prior

Ali Abbasi and Mohammad Rahmati, Member, IEEE

Abstract—Over the past few decades, numerous attempts have been made to address the problem of recovering a high-resolution (HR) facial image from its corresponding low-resolution (LR) counterpart, a task commonly referred to as face hallucination. Despite the impressive performance achieved by position-patch and deep learning-based methods, most of these techniques are still unable to recover identity-specific features of faces. The former group of algorithms often produces blurry and oversmoothed outputs particularly in the presence of higher levels of degradation, whereas the latter generates faces which sometimes by no means resemble the individuals in the input images. In this paper, a novel face super-resolution approach will be introduced, in which the hallucinated face is forced to lie in a subspace spanned by the available training faces. Therefore, in contrast to the majority of existing face hallucination techniques and thanks to this face subspace prior, the reconstruction is performed in favor of recovering person-specific facial features, rather than merely increasing image quantitative scores. Furthermore, inspired by recent advances in the area of 3D face reconstruction, an efficient 3D dictionary alignment scheme is also presented, through which the algorithm becomes capable of dealing with low-resolution faces taken in uncontrolled conditions. In extensive experiments carried out on several well-known face datasets, the proposed algorithm shows remarkable performance by generating detailed and close to ground truth results which outperform the state-of-the-art face hallucination algorithms by significant margins both in quantitative and qualitative evaluations.

Index Terms—Image Super-Resolution, Face Hallucination, Sparse Representation, Face Subspace, 3D Dictionary Alignment.

I. INTRODUCTION

Our desire to enhance the resolution of an already-recorded image is arguably as old as the time when the early photographs were taken. With the emergence of digital images, the idea also started to attract the attention of many researchers, leading to the introduction of a popular field in the area of image processing, known as image super-resolution [1]. Even today, despite cameras having ever-increasing resolution, there is still a huge demand in increasing the resolution of the existing images, particularly in specific applications such as law enforcement, surveillance, and monitoring, where images are taken under uncontrolled conditions and are required to be further processed before being used for a particular purpose. More importantly, most computer vision algorithms are designed to work with high quality images, which means their performance would severely affected when given a low-resolution input [2].

Among different variations of image super-resolution applications, those which deal with super-resolving face images have always been of special interest to researchers, and are often categorized under the name face hallucination. The term was first coined by Baker and Kanade [3] in 2000, and since then has gained huge popularity due to its wide range of applications, with dozens of algorithms proposed so far.

One can hardly offer an explicit classification of the algorithms presented in the literature, as in many cases, the distinction among different categories of methods is not clear. Consequently, different studies have suggested different criteria to classify face hallucination algorithms; including operating domain (spatial vs. frequency), number of input images (single vs. multiple), and reconstruction method (reconstruction-based vs. learning-based). Early researchers relied more on statistical approaches to predict the HR face image given the LR observation. In their pioneering effort [3], Baker and Kanade used a Bayesian framework with gradient priors to estimate high-frequency components of a face image. Inspired by their work, Su et al. [4] proposed a similar formulation in which the prior was estimated by matching local low-level facial features from the input LR and the training HR face images. Meanwhile, Markov random fields (MRF) also started to draw the attention of researchers, after a two-step method was suggested by Liu et al. [5].

Another major group of researchers focused on making use of training samples to learn a projection matrix which could be later used to project the LR input into high dimensional space and obtain the reconstructed HR output. They based their idea on the fact that face images share structural similarities, and therefore can be synthesized from a linear combination of other samples. Wang and Tang [6] addressed the problem by applying PCA to fit the input face image as a linear combination of the training low-resolution face images, and then reconstructing the HR output by using the combination weights for the training high-resolution images. Despite considerable performance, their method failed to recover fine details of face images as it only focused on global face information. To alleviate this, various methods has been suggested in the literature. In [7], authors adopted locality preserving projection (LPP) to learn the projection weights. Yang et al. [8] employed non-negative matrix factorization (NMF) to find the face subspace, along with a patch-based sparse representation method using coupled overcomplete dictionaries to generate final hallucinated image. Also in [9], the coefficient vector was obtained through a recursive error back-projection method.

To find the aforementioned subspace, many studies have utilized the idea of manifold learning by assuming that LR face images and their HR counterparts are sampled from two
manifolds which have similar local neighborhood structures. Liu et al. [10] maintained the local features by developing a multilinear patch-based reconstruction method. Fan and Yeung [11] addressed the problem through a two-step approach using neighbor embedding over visual primitive features. Huang et al. [12] applied canonical correlation analysis (CCA) to determine this subspace. Authors in [13] managed to learn pixel-wise structure prior represented as embedding coefficients to estimate the final result. Because of the one-to-many mapping relation between LR and HR samples, some researchers cast doubt on the above manifold assumption and suggested different alternative strategies. Li et al. [14] presented a manifold alignment approach which projected the two manifolds to a common hidden manifold. In another study [15] a strategy was devised to learn linear models based on the local geometrical structure on the high-resolution manifold. To avoid dealing with the difficulties of preserving local geometry in various resolutions, [16] directly regularized the relationship between target patch and training patches in the HR space. Later, Shi et al. [17] addressed this challenge by training a series of adaptive kernel regression mappings for predicting the missing details from LR patches.

Position-patch based face hallucination methods have also gained wide popularity during the last decade. The main intuition behind these algorithms is that the HR counterpart of a given input LR image patch can be reconstructed by applying neighbor embedding to those patches located in the same position as the test patch. Ma et al. [18] was first to suggest this method, by computing the reconstruction weights through solving a least square problem. To obtain a more suitable solution, Jung et al. [19] decided to replace the least square estimation with a convex constrained optimization. Various attempts have been made recently to use the idea of locality-constrained representation (LcR) in order to impose a locality constraint on the least square inversion problem to encourage sparsity and locality simultaneously [20]–[22].

In recent years, with the advancement of neural networks, deep learning-based face hallucination algorithms have become increasingly prevalent in the literature. Motivated by powerful representation abilities of CNNs, Zhou et al. [23] designed a network architecture to learn the mapping between the raw input image and the face representations extracted by a deep convolutional network. To avoid oversmoothing problem and preserve more textural details, WaveletSRNet [24] reconstructed HR images in wavelet coefficient domain. Chen et al. [25], extracted multi-scale features by incorporating multiple encoders and decoders in bottom-up and top-bottom patterns. In [26], a super-resolution technique was suggested which decomposed faces and recovered different components. Jiang et al. [27] also developed a network with two individual branches to learn global facial shape and local facial components.

The emergence of generative adversarial networks had also a great impact on face super-resolution studies. Yu et al. [28] pioneered in GAN-based face hallucination algorithms, by considering an architecture which consisted of two discriminative and generative networks. They further extended their work [29] by incorporating multiple spatial transformer networks (STN) in their network to improve the alignment and upsampling performance. Later, they enhance robustness against noisy inputs and inputs with non-fixed resolution in [30] and [31], respectively. Bulat et al. [32] discussed the idea of learning the degradation before super-resolution in a two-stage process. In [33], the problem was formulated with a collaborative suppression and replenishment framework, whereas the algorithm of [34] made the training phase more effective and efficient by introducing spatial attention into the generator. A self-supervised method in which the problem of face super-resolution is expressed as generation problem was developed in [35], [36] also took a different approach by integrating multiple deep learning networks of different types.

A. Motivation and Contribution

Recent face hallucination studies have been dominated by two approaches: position-patch and deep learning-based methods. The first relies on the basic assumption that small patches in LR and HR spaces create manifolds with similar local geometry, hence they consider the reconstruction weights in both spaces equal. However, it has been shown [14] that, due to the nonisometric one-to-multiple mappings from LR patches to HR ones, this assumption is not always met in practice. Therefore, face images of two entirely different individuals may have similar LR patches, whereas HR/LR patch pairs of a specific person may bear no similarity at all [37]. This becomes more severe as the LR input degradation level increases [21], or patches with smaller sizes are considered. Fig. 1 illustrates four patches of different sizes extracted from the same LR image. As shown, training patches belonging to the test subject tend to be more similar to the LR test patch when they increase in size. To further demonstrate this point, the average neighborhood preservation rates (NPR) [38] between the LR and HR image manifolds according to different patch size and based on three levels of degradation is presented in Fig. 2. As the graph suggests, selecting patches with bigger sizes gradually increases the average NPR, hence the reliability.
of the selected patches. However, the rising trend gradually weakens before the case when the whole image is taken into consideration, in which a sudden jump in the values of NPR is observable. The figure also reveals that, as image becomes more degraded, more invalid patches will be selected as the neighboring patches. Still, the case when the entire image is considered is relatively less affected by this change.

Generally, patch-based methods often face the dilemma of selecting appropriate patch size. On one hand, to capture the global nature of faces and finding more meaningful and accurate neighbors, larger patches are preferred. On the other hand, as a consequence of the curse of dimensionality [39], the size of the training set should grow exponentially with the patch size to guarantee valid matches [40] and avoid ghosting effects [22]. Although several approaches have been suggested to alleviate this problem [21], [22], [37], existing patch-based methods still fail to recover person-specific facial features due to the above-mentioned limitations.

A similar argument can also be made in the case of deep learning-based face super-resolution algorithms. In spite of their great ability to add visually pleasing details to LR images, these algorithms often neglect how much beneficial the added information is for the task of recognizing the identity of the face [41]. Most of the loss functions that have been considered in the literature are designed to minimize the mean square error (MSE) between the HR image and its corresponding reconstructed one, which, although can sometimes achieve high MSE-oriented quality metrics, in most cases produce blurry and over-smoothed results [42]. In order for deep learning-based methods to be able to learn identity-aware representations, they are required to be trained with a large well-labeled face dataset, which tends to be very costly [41]. In recent years, several network architectures and loss functions have been suggested to incorporate the identity prior into the learning procedure [41], [42]. However, still in many cases the hallucinated face hardly resembles the person in the test image, as shown in Fig. 3.

In this paper, a novel face hallucination approach is presented in which super-resolution is performed in the subspace spanned by the available training faces, often referred to as

**face subspace** [43]. To accomplish this, face subspace prior has been incorporated as a regularization term, and through a simple, yet vastly effective MAP-based formulation, the benefits of global hallucination are achieved whereas the drawbacks of patch-based methods are avoided. Additionally, although the proposed algorithm can be considered as a global reconstruction scheme, however, the hallucinated faces are artifact-free and robust to ghosting effects, and there is no constraints on dataset size either. The optimization process of the proposed objective function is also addressed through a highly effective recently introduced closed-form solution [44].

Furthermore, to better deal with the cases where there is a significant misalignment between the input LR face and the training faces, an effective 3D dictionary alignment technique which allows us to perform face super-resolution on unconstrained face images will be suggested. The proposed alignment procedure can also be used in the pipeline of other similar algorithms, and since the majority of current face hallucination approaches only produce satisfactory results when given frontal LR face images, this method can substantially increase their robustness against pose variations in LR inputs.

Therefore, the major contributions of the paper can be summarized as follows:

- We force the reconstructed face to lie in the linear span of the training faces, hence, unlike most existing face hallucination algorithms, the reconstruction is performed in favor of both recovering identity-specific face attributes as well as enhancing image quantitative measures.
- We will show that not more than three samples per subject are required for our algorithm to guarantee an identity-preserving result and outperform the existing methods in both tasks of face super-resolution and face recognition.
- By incorporating an efficient 3D alignment procedure, the algorithm can extend its superior performance to the case where LR face pose is significantly different from the ones in the training set. More importantly, the proposed alignment scheme allows us to deal with face hallucination problems in which face images from both the training and testing sets are unconstrained.
- In contrast to the patch-based algorithms, the proposed method is barely affected by increasing the level of degradation, and shows outstanding robustness when given very low-resolution (VLR) face images.
- By utilizing an efficient closed-form solution for the proposed objective function, our method is considered to be a very fast face hallucination algorithm in which the computational time is comparatively less affected by increasing the size of the HR image or the number of samples in the dataset.
- The proposed method achieves superior performance over the competitive state-of-the-art algorithms in various frontal, non-frontal, and in-the-wild face hallucination experiments conducted on several well-known face datasets, and surpasses the position-patch and deep learning-based methods both quantitatively and qualitatively.

The organization of the rest of the paper is as follows: in section II, the face subspace prior along with our main MAP-
A common assumption in the problem of single image super-resolution is that the low-resolution input is a noisy, blurred, and decimated counterpart of the unknown high-resolution image. Consequently, the following forward degradation model is often taken into consideration:

$$y = SHx + \epsilon$$  \hspace{1cm} (1)

in which \(y \in \mathbb{R}^{m_l \times 1}\) is the low-resolution input with \(x \in \mathbb{R}^{m_h \times 1}\) as its high-resolution equivalent, where \(m_l = w_l \times h_l\) and \(m_h = w_h \times h_h\). In addition, \(H \in \mathbb{R}^{m_h \times m_h}\) represents blurring filter, \(S \in \mathbb{R}^{m_l \times m_h}\) denotes decimation operator with scaling factor \(d\), hence \(m_h = m_l \times d^2\), and \(\epsilon \in \mathbb{R}^{m_l \times 1}\) indicates additive white Gaussian noise (AWGN) encountered through the image acquisition process. The problem of super-resolution can therefore be written as an optimization problem derived from maximum likelihood (ML) estimator of the high-resolution image \(x\) as below:

$$\hat{x} = \arg \min_x \|y - SHx\|^2_2$$  \hspace{1cm} (2)

which leads to the following solution:

$$\hat{x} = (H^T S^T S H)^{-1} H^T S^T y$$  \hspace{1cm} (3)

This solution, which is equivalent to the least-square solution of the inverse problem of (1), is known to be an ill-conditioned problem due to its sensitivity to small noise and measurement errors [1]. On condition that \(H^T S^T S H\) is singular, the problem is also ill-posed with infinite space of possible solutions. Moreover, solving (3) requires inverting the matrix \(H^T S^T S H\) with a computational complexity of the order \(O(m_h^3)\), which makes it practically inefficient in many real scenarios [44].

To overcome these problems, some additional information is needed to constrain the space of solutions and stabilize the problem. This is often accomplished by introducing a new term to (2), converting the maximum likelihood problem to a maximum a posteriori (MAP) problem:

$$\hat{x} = \arg \min_x \|y - SHx\|^2_2 + \xi \Phi(Rx)$$  \hspace{1cm} (4)

which consists of a fidelity term corresponding to model likelihood and a regularization term which represents a priori knowledge about the original image, with the regularization parameter \(\xi\) which determines the contribution of each term. \(R\) is also a matrix which can be defined according to the application. Various priors for natural images have been suggested in the literature, with Tikhonov regularization and Total Variation as the most notable ones [1]. However, to the best of our knowledge, there are few, if any, such priors introduced for the purpose of addressing single-frame global face hallucination problem, and the studies mostly include multi-image [45] or patch-based [46] approaches.

A. Identity-Preserving Face Prior

In the area of pattern recognition, a well-established assumption is that patterns from a specific object class lie on a linear subspace [59]. In regard to facial recognition problem, it has been verified that face images belonging to a certain subject create a low-dimensional subspace, and the idea has been the cornerstone of various successful face recognition algorithms [43]. Let \(D_h = [x_{i,1}, x_{i,2}, \cdots, x_{i,n_i}]\) be the HR training faces of the \(i\)th subject. On condition that \(n_i\) is sufficiently large, the above assumption implies that if the hallucinated face image \(x\) belongs to the subject \(i\), for some scalar coefficients \(\alpha_{i,j} \in \mathbb{R}\), \(j = 1, 2, \ldots, n_i\), it can be represented as \(x = \alpha_{i,1} x_{i,1} + \alpha_{i,2} x_{i,2} + \cdots + \alpha_{i,n_i} x_{i,n_i}\). Therefore, in case the subject of the input LR face is given beforehand, a suitable prior term for (4) would be

$$\Phi(x) = \|x - \alpha_{i,1} x_{i,1} - \alpha_{i,2} x_{i,2} - \cdots - \alpha_{i,n_i} x_{i,n_i}\|^2_2$$  \hspace{1cm} (5)

To generalize the above term, a dictionary matrix \(D_h\) is defined which contains the whole \(n\) training faces of all \(c\) subjects, in which, to facilitate classification task, face images of the same subject are arranged beside each other, that is, \(D_h = [D_{h,1}, D_{h,2}, \cdots, D_{h,c}] = [x_{1,1}, x_{1,2}, \cdots, x_{c,n_c}]\). Therefore, (5) changes to

$$\Phi(x) = \|x - D_h \alpha\|^2_2$$  \hspace{1cm} (6)

where \(\alpha\) is the coefficient vector with non-zero entries for those elements associated with subject \(i\), and zero elsewhere. Incorporating this prior term into (4) gives

$$\{\hat{x}, \hat{\alpha}\} = \arg \min_{x, \alpha} \|y - SHx\|^2_2 + \mu \|x - D_h \alpha\|^2_2 + \lambda \|\alpha\|_1$$  \hspace{1cm} (7)

Provided that the number of subjects is sufficiently large, \(\alpha\) is expected to have a sparse representation, since only few of its elements will have nonzero values. Therefore, a new regularization term \(\|\alpha\|_0\) should be included in (7), and since this leads to an NP-hard problem, the \(l_0\)-norm is replaced by \(l_1\)-norm [43]:

$$\{\hat{x}, \hat{\alpha}\} = \arg \min_{x, \alpha} \|y - SHx\|^2_2 + \mu \|x - D_h \alpha\|^2_2 + \lambda \|\alpha\|_1$$  \hspace{1cm} (8)

With the above formulation, the hallucination is constrained to the subspace spanned by the subject which gives the sparsest coefficient vector with respect to the input image. Thus, if there are enough samples of the subject to which the input face belongs, the super-resolution is performed for the benefit of face recognition. Later, the details of the optimization procedure of the proposed algorithm is scrutinized, before the introduction of our 3D dictionary alignment pipeline which is detailed in the same section. The experimental evaluations and comparison with other competitive algorithms are the subject of section III, followed by conclusion and possible future works which are presented in section IV.
of recovering true facial attributes. In section [III] we will show that not more than three samples per subject are required to satisfy this condition.

B. Optimization

The optimization problem (3) can be divided into two subproblems associated with each of the variables \( x \) and \( \alpha \), before being solved iteratively for one while fixing the other. The following two optimization steps can therefore be defined:

1) Optimizing for \( x \): The intermediate HR estimate of the LR input \( y \) in a given iteration \( t \) can be obtained through the following \( l_2 \)-regularized optimization problem:

\[
x_{t+1} = \arg \min_x \|y - S H x\|^2_2 + \mu \|x - D_h \alpha_t\|^2_2
\]

(9)
whose closed-form solution is given by

\[
x_{t+1} = (H^H S^H S H + 2 \mu I_{m_h})^{-1}(H^H S^H y + 2 \mu D_h \alpha_t)
\]

(10)

Unlike the optimization procedure of other similar inverse problems (e.g., image deblurring [47]) which can be solved efficiently in the frequency domain, here, the presence of the decimation operator \( S \) in the fidelity term, and the fact that the product matrix \( S H \) does not have a block-circulant structure and cannot be diagonalized in the frequency domain, makes the problem impossible to be solved using the Fourier transform. However, [44] showed that under certain assumptions on the decimation operator \( S \) and the blurring matrix \( H \), the optimization problem admits a closed-form solution in the frequency domain.

More precisely, assuming \( H \) as the matrix representation of the cyclic convolution operator, one can decompose the blurring operator \( H \) and its conjugate transpose \( H^H \) as

\[H = F H \Lambda F, \quad H^H = F H^H \Lambda^H F^H\]

(11)

where \( F \in \mathbb{C}^{m_h \times m_h} \) is the discrete Fourier transform matrix with the property \( F^H = F^{-1} \), hence \( F F^H = F H^H F = I_{m_h} \), and \( \Lambda \in \mathbb{C}^{m_h \times m_h} \) is a diagonal matrix whose elements are the Fourier transform of the zero-padded PSF, that is, the first column of the blurring matrix \( H \).

Additionally, \( S \) is assumed to be a downsampling operator whose conjugate transpose \( S^H \) interpolates the decimated image with zeros, and satisfies the relationship \( S S^H = I_{m_l} \).

By considering \( S^H \otimes S \), which operates as an element-wise multiplication by an \( m_h \times m_h \) matrix with ones at the sampled positions and zeros elsewhere, [48] showed that:

\[F S F^H = \frac{1}{d^2} J_{d^2} \otimes I_{m_l}\]

(12)
in which \( \otimes \) denotes the Kronecker product, \( J_{d^2} \in \mathbb{R}^{d^2 \times d^2} \) is a matrix of ones, and \( I_{m_l} \in \mathbb{R}^{m_l \times m_l} \) is an identity matrix. Bearing in mind [12] as well as the previously mentioned assumptions, one can rewrite the analytical solution (10) as

\[x_{t+1} = F H \left( \frac{1}{d^2} \Lambda^H \Lambda + 2 \mu I_{d^2} \right)^{-1} F (H^H S^H y + 2 \mu D_h \alpha_t)
\]

(13)

where \( \Lambda \in \mathbb{C}^{m_l \times m_h} \) is defined as \( \Lambda = [\Lambda_1, \Lambda_2, \ldots, \Lambda_d^2] \), with the blocks \( \Lambda_i \in \mathbb{C}^{m_l \times m_l} (i = 1, \ldots, d^2) \) such that \( \text{diag}\{\Lambda_1, \ldots, \Lambda_d^2\} = \Lambda \). This can be further simplified by incorporating the Woodbury matrix identity [49] into (13) and obtaining the following closed-form solution:

\[x_{t+1} = \frac{1}{2 \mu} \left( r - F H^H \Lambda \left( 2 \mu d^2 I_{m_h} + \Lambda \Lambda^H \right)^{-1} F^H y \right)
\]

(14)

where \( r = H^H S^H y + 2 \mu D_h \alpha_t \).

2) Optimizing for \( \alpha \): The coefficient vector \( \alpha \) is updated using the following well-known \( l_1 \)-minimization problem:

\[\alpha_{t+1} = \arg \min_{\alpha} \|x_t - D_h \alpha\|^2_2 + \lambda \|\alpha\|_1
\]

(15)

which can be efficiently solved using various \( l_1 \)-minimization algorithms. It should also be noted that the initial \( \alpha \) vector is obtained through solving the above minimization problem for the input face \( y \) and the low-resolution training dictionary \( D_l \), i.e., \( \alpha_0 = \arg \min_{\alpha} \|y - D_l \alpha\|^2_2 + \lambda \|\alpha\|_1 \).

Algorithm [1] summarizes the entire procedure of the proposed face hallucination approach.

C. 3D Dictionary Alignment

As face images of different poses are distributed on a highly nonlinear manifold [50], the majority of dictionary-based face hallucination algorithms fail to achieve satisfactory results when given non-frontal input faces. Inspired by recent advances in 3D face reconstruction and alignment studies, in this section an efficient dictionary alignment procedure will be presented, by which the training faces are registered with respect to the LR face pose before being used in the hallucination process. This additional step gives significant flexibility to the main algorithm and boosts its performance in reconstructing LR faces in the presence of high pose variations, even when the training faces are also non-frontal and unconstrained. The proposed alignment procedure, which is visually presented in Fig. [4], contains the following steps:

1) Training Faces 3D Reconstruction: In order to perform 3D alignment, we first use [51] to generate 3D geometries associated with each of the training faces, and obtain two sets \( \{M_i\} \) and \( \{C_i\} \), such that \( M_i = (V_i, T_i) \) is the 3D mesh associated with the \( i \)-th training face, with \( V_i \in \mathbb{R}^{n_v \times 3} \) and \( T_i \in \mathbb{R}^{n_t \times 3} \), and \( C_i \in \mathbb{R}^{n_c \times 3} \) as its vertices, triangles, and color attributes, where \( n_v \) and \( n_t \) denote the number of vertices and triangles, respectively. As might be expected, the whole
process of face reconstruction is performed offline, hence does not affect the runtime of the main hallucination process.

2) LR Face Landmark Detection: The most crucial part of the alignment pipeline is to locate facial landmarks on the LR input face. Since the proposed hallucination algorithm will accept degraded facial images with high variations in pose, the landmark detection method is expected to be highly robust and perform well in uncontrolled conditions. Fortunately, recent advances in deep neural networks has allowed researchers to propose powerful facial landmark detectors with considerable speed, accuracy, and stability. According to our experiments, most of the current state-of-the-art approaches fully satisfy our desired level of robustness, and therefore are eligible to be used in the alignment procedure. Fig. 5 illustrates the average normalized mean error (NME) between the landmarks of a set of upscaled LR face images detected by [52] and their corresponding ground truth points, based on different levels of degradation. According to the figure, in most cases the difference between landmarks detected in the LR face images and their ground truth points is fairly negligible, even when face images as small as $15 \times 15$ pixels with $5 \times 5$ blur kernel are considered. Since conventional degradation settings are often far more lenient, we can rest assured that the detected landmarks $p_y$ will not affect the main alignment procedure.

3) Registering Training Faces: Having the 3D landmark points associated with both the training HR faces and the input LR face, one can easily estimate the transformation matrices needed for registering each training face with respect to the LR input face. To speed up the process, a set of landmark points $p_{x,f}$ are considered as the reference, and the transformation matrices $\{\tau_{p_1,p_{x,f}}, \tau_{p_2,p_{x,f}}, \ldots, \tau_{p_n,p_{x,f}}\}$ are calculated in the training phase, in which $\tau_{p_i,p_{x,f}}$ denotes the transformation between the landmarks associated with the $i$th training sample and the reference. In the test phase, it is only required to find the transformation matrix $\tau_{p_{ref},y}$ between the LR input face and the reference landmarks, and obtain $\{\tau_{p_1,y}, \tau_{p_2,y}, \ldots, \tau_{p_n,y}\}$ simply by $\tau_{p_i,y} = \tau_{p_i,p_{x,f}} \cdot \tau_{p_{x,f},y}$. Performing the registrations and finding the transformed face objects is also efficiently implemented by applying each transformation to the corresponding object vertices.

4) 3D Face Rendering: Finally, the 3D face objects are converted to 2D images using the available object rendering algorithms. This step is the most time-consuming part of the whole process, and extra care must be taken to preserve the details and information of the face object.

The above steps are performed on each of the training samples to obtain the aligned HR and LR dictionaries $D_h^a$ and $D_f^a$. To reduce the error caused by non-face regions in the process of hallucination, we also remove the excess pixels in the LR input face using a mask extracted from the average of the registered LR faces to obtain $y''$, which will be later used along with $D_h^a$ and $D_f^a$ in Algorithm 1 to perform pose-robust face hallucination. As will be shown in section III the proposed 3D dictionary alignment significantly improves the face hallucination performance even in the presence of high pose variations. Moreover, the whole process takes roughly...
0.04 seconds for a 20 × 20 training face image and its corresponding object with $n_e = 43,867$ and $n_a = 86,906$, which is indeed a reasonable time considering the amazing benefits it offers. A summary of the 3D dictionary alignment procedure is presented in Algorithm 2 as pseudo-code.

Algorithm 2 3D Dictionary Alignment

**Input:** Training face objects $\{M_i\}_{i=1}^n$ and their color attributes $\{C_i\}_{i=1}^n$, reference landmark points $p_{ref}$, transformations between the training faces and the reference landmarks $\{\tau_{p_1,p_{ref}}, \tau_{p_2,p_{ref}}, \ldots, \tau_{p_n,p_{ref}}\}$, low-resolution image $y$.

1. $y_{bic} \leftarrow$ upscale$(y)$
2. Extract 3D facial landmark points from $y_{bic}$.
3. Align $y_{bic}$ with respect to the reference landmarks $p_{ref}$ and obtain the transformation matrix $\tau_{p_{ref},p}$.
4. for $i = 1 : n$
5. \[ \tau_{p_i,p} = \tau_{p_{ref},p} \]
6. \[ V_i^T = V_{ref} \]
7. \[ x_i \leftarrow \text{render}(V_i^T, T_i, C_i) \]
8. Add $x_i$ to the HR dictionary $D^h_i$.
9. end for
10. $D^o_n \leftarrow \text{degrade}(D^o_k)$
11. Obtain $y^m$ by applying the mask extracted from the average of the LR aligned faces to $y$.

**Output:** Aligned dictionaries $D^h_n$ and $D^o_n$, masked LR input $y^m$.

### III. Experimental Results

In this section, several experiments have been carried out to evaluate the performance of the proposed algorithm and demonstrate its efficiency and applicability. For this purpose, a number of recently published state-of-the-art face hallucination methods have been selected with their parameters tuned so that they produce their optimal results. The experiments on frontal face hallucination are performed on the FERET [53], the CMU Multi-PIE [54], and the AR [55] public face datasets, whereas the pose-robust face super-resolution algorithm is tested on the CMU Multi-PIE and the LFW [56] databases.

#### A. Experiments on Frontal Faces

The proposed method is first assessed on faces taken in controlled condition. Unless otherwise specified, all the LR face images are obtained after applying downsampling and blurring (by a $4 \times 4$ average smoothing filter) to their HR counterparts. Samples from all databases were aligned based on the location of the eye corners. For all the experiments, one random image per each subject was selected as the test sample and the remaining were used in the training phase. The regularization parameters $\mu$ and $\lambda$ are chosen to be $10^{-8}$ and $2700$, respectively, whereas the number of iterations $T$ is set to 30. In the patch-based approaches, the patch size and the overlapping parameters are chosen according to the LR and HR image sizes. In [6], the eigenvalues accumulation contribution rate is set to 0.99. [15] is modified so that it includes the blur information of the LR inputs. The implementations of [20] and [21] were slightly changed to prevent errors in recovering very low-resolution face images.

1 All the implementations used in the evaluations are from the official source codes released by the authors.

1) The FERET dataset: We first evaluate the performance of the proposed method on the frontal facial images from the FERET database. We only select subjects with five or more samples in the database, which leads to a subset of 519 images from 70 individuals, each with unequal number of samples. The LR input faces are of size 15 × 10, and scaling factors are set to 2, 4, and 8. The PSNR and SSIM performance of different methods is summarized in Table I. In all three experiments and with different scaling factors, the proposed method outperforms the second best algorithm by 1.18 dB, 1.00 dB, and 1.00 dB in PSNR and 0.0095, 0.0199, and 0.0353 in SSIM, respectively. In Fig. 7 the performance of the algorithms on each one of the test samples when scaling factor is 4 is displayed, which demonstrates the dominance of the proposed method over the competitive ones in almost all the available test images. Fig. 6 (top three rows) also qualitatively compares the methods on three testing images with different scaling factors. It is observable that the competitive methods were unable to recover facial details including eyeglasses and wrinkles. In the presence of facial expressions, [6] produced undesirable artifacts in the recovered face images. The methods based on position-patch also produced blurry and oversmoothed images, particularly around the mouth regions. The results of LM-CSS [15] appear to be more similar to the original HR faces than those of the remaining approaches, however, the noise and artifacts added to the resultant faces have made this method quantitatively unsatisfactory. In general, the results obtained by the proposed algorithm are obviously more detailed, clear, and artifact-free compared to those produced by the others.

2) The Multi-PIE dataset: We next focus on the Multi-PIE database frontal images (camera 05-1) taken in neutral expression and under normal illumination (illumination condition 10) from all four sessions and two recordings. Hence there will be 650 images representing 130 subjects, each with five different samples. In this experiment, the size of the LR input images are $12 \times 9$, and three different scaling factors 4, 8, and 16 are considered. The quantitative results of different
methods can be seen in Table II. One can notice that the proposed method shows superior performance over the other algorithms by considerable margins (1.09 dB, 1.46 dB, and 1.28 dB in PSNR and 0.0167, 0.0445, and 0.0308 in SSIM, in scaling factors 4, 8, and 16, respectively). Visually speaking, as depicted in the second three rows of Fig. 6, the results of the proposed algorithm is by far more similar to the ground truth images, particularly when individual-specific face attributes (such as hair, mouth, wrinkles, and eyeglasses) are taken into consideration. Due to their inability to recover edges, almost all the other methods were unable to reconstruct age-specific features of the faces, leading to faces which appear younger than the person’s actual age. Owing to its locality constrained approach, TLcR-LR [22] has done slightly better than the other
methods in recovering some facial attributes (e.g., eyeglasses), however, its results still suffer from blurriness and lack of details.

3) The AR dataset (VLR Face Hallucination): To demonstrate the efficiency of the proposed algorithm in hallucinating very low-resolution face images [2], extensive experiments were conducted on the AR face database. From the cropped version of the database [59], we discard the images with occlusion, and select a subset of 1400 images associated with each of the 100 subjects, each with 14 samples. Low-resolution input faces are chosen to be 5 × 4, making it an extreme case of VLR face super-resolution task with only 20 LR pixels available. The dataset also contains facial images with significant expression variations, which cause the problem to be even more challenging. We upscale the existing LR faces by the factors of 4, 8, and 16. As can be seen in Table III, our approach shows its capability in recovering very low-resolution inputs in all three experiments by improving the PSNR by 1.18 dB, 2.37 dB, and 1.5 dB, and the SSIM by 0.0186, 0.1235, and 0.1166 units, respectively. More importantly, as the visual comparison in the last three rows of Fig. 6 suggests, the performance of the position-patch based methods falls dramatically when given VLR inputs, with their results being blurry and mostly irrelevant. Conversely, despite being a classic algorithm, [6] manages to outperform several recently introduced patch-based techniques due to its global reconstruction approach. All in all, the results achieved by the proposed algorithm bear much more visual resemblance to the ground truth faces compared to those of the others.

B. Parameters Analysis

1) Regularization Parameters: There are two regularization parameters in the proposed formulation, namely \( \mu \) and \( \lambda \). The first determines the closeness of the reconstructed face to the subspace spanned by the available faces, whereas the second decides how strictly this face subspace should be estimated. In this subsection, we perform experiments on a randomly selected face image and tune each parameter separately while keeping the other fixed. By fixing \( \mu \) and changing \( \lambda \) values over the range \([0, 10^4]\) with an interval of 500, as plotted in Fig. 8, one can notice that the best performance is achieved when \( \lambda \) is roughly set to \( 10^3 \). We next fix the value of \( \lambda \) and chose 20 different values for \( \mu \) from the range \([0, 20]\). The variations of PSNR and SSIM (Fig. 9) suggest that values closer to zero are more desirable for this parameter. When \( \mu \) is set to zero, however, the hallucinated face image will be incalculable (NaN) due to the absence of regularization term in the optimization function. We therefore set \( \lambda = 2700 \) and \( \mu = 10^{-8} \) in our experiments.

2) Number of Training Samples per Subject: The proposed algorithm utilizes the face subspace as a prior knowledge, thus it is desirable to see how the quality of the subspace spanned by the training face images affects its performance. In this regard, the diversity among the faces (i.e., the availability of faces with different variations), which is related to the number of training faces per subject, is expected to be decisive. To justify this, we create a test set by randomly selecting one sample from each subject of the AR dataset, and perform a series of experiments on the selected set, each time with different number of training samples per subject. Fig. 10 shows the quantitative metrics obtained by different methods in each experiment. It is observable that when there is only one sample per subject, TLcR-RL [22] and TRNR [58] achieve better performance compared to the proposed method. This can
be justified by the fact that in this case, no subject-specific face subspace is formed and subsequently the prior term used in the formulation will be of no benefit. However, the performance of the proposed method improves significantly when another training sample is added for each subject, with only 0.1 dB lower PSNR and 0.0246 higher SSIM compared to the pioneering method. When there are three samples per each subject, the proposed method clearly outperforms the remaining algorithms, and by increasing the number of samples continues to improve its dominance whereas the performance of the other methods remains relatively unchanged. In Fig. 11, the influence of the number of samples per subject on the visual appearance of a face image reconstructed by the proposed method (bottom) and [22] (top) is displayed. As the number of samples per subject increases, more details appear in the hallucinated face, and the undesired effects (for example, on forehead region) diminish. With four samples per subject, a clean face image close to the ground truth is obtained, whereas the true facial expression is reconstructed when seven samples per subject are used. The figure also reveals that the results generated by TLcR-RL are less affected by the addition of extra samples to the training set, and are still oversmoothed and blurry even when there are 13 samples per subject available. To summarize, the experiment demonstrates that the proposed algorithm requires only two to three images per each subject to outperform the competitive methods both quantitatively and qualitatively.

3) Number of Iterations: To investigate the influence of the face subspace prior, it would be worthwhile to compare the average PSNR and SSIM values across different iterations. As shown in Fig. 12, the two quantitative metrics improve dramatically in the early iterations, reaching to 32.92 dB in PSNR and 0.9622 in SSIM in a single iteration (2.65 dB and 0.0278 improvements, respectively), and surpassing 33.64 dB in PSNR and 0.9670 in SSIM after the first five iterations (3.36 dB and 0.0326 improvements, respectively), with respect to the initial state, i.e., $\hat{x} \approx D_0 \alpha_0$. The growth of these two indicators becomes stable roughly at iteration 30, thus, it will be considered as the number of iterations in our experiments.

4) Blur Kernel Size: We further test the performance of our approach against various levels of degradation to measure how robust the proposed algorithm is when less facial information is available. We set the LR input size and the upsampling factor to be $12 \times 9$ and $4$, respectively, and perform several experiments with average blur kernel of different sizes. The quantitative results of different approaches in this experiment are represented in Fig. 13. Despite the significant decline in the performance of the other methods, our algorithm is barely affected by the increase in blur kernel size, and even in extreme cases, its quantitative measures remain more or less the same. As discussed in section 4, as a result of selecting inappropriate neighboring patches, the performance of the patch-based face hallucination methods is prone to be seriously affected when LR input images become more degraded.
C. Influence of the Face Subspace Prior

The effectiveness of the face subspace prior in the proposed approach is further evaluated by making more quantitative and qualitative comparisons between the final reconstructed face obtained by our algorithm and its initial state, which is the case when only the concept of neighbor embedding is taken for granted and the hallucinated face is equal to $D_k \alpha_0$. Fig. 14 displays the quantitative indicators for the test faces in the FERET database. The linear embedding-based approach achieves the average 30.29 dB in PSNR and 0.9344 in SSIM, which are lower than the proposed algorithm by 3.42 dB and 0.0323, respectively. This substantial difference between the initial and final states of the algorithm is the result of the improvements made by the face subspace prior, which has also been clearly reflected in the facial details recovered by the two approaches. According to the examples shown in Fig. 15 in the initial phase of the algorithm, only basic structures of the faces are recovered, and, in contrast to the final hallucination results, fine details such as eyeglasses, nose shape, eyebrows direction, and face pose are all ignored. As a consequence of considering the faces merely as the linear combinations of the training samples, these images contain blurry regions and in some cases are by no means close to their ground truth counterparts.

D. Face Recognition Accuracy

To clarify the advantages of our algorithm in recovering person-specific facial features, we conduct experiments on the task of low-resolution face recognition. The evaluations are performed on the Multi-PIE and the AR face datasets with 130 and 100 subjects, respectively, such that each subject has only two samples in the training set. The size of the LR images in the Multi-PIE subset is $8 \times 6$, whereas the inputs in the AR subset are of size $10 \times 8$, and the scaling factor in both experiments is 4. LR faces are obtained by applying a $7 \times 7$ Gaussian filter with $\sigma = 2$ to the HR images, before downsampling them to the desired sizes. The resultant images of all methods are classified using the SRC classifier [43]. As for the proposed method, the final coefficient vector $\alpha$ is used in the classification. Table IV represents the recognition accuracy achieved by different approaches in both experiments, whereas Fig. 16 compares their cumulative recognition rates in the first five ranks. The recognition rates of the proposed method on the Multi-PIE and the AR datasets are 92.25% and 76%, outperforming the others by 8.53% and 3%, respectively. This clearly illustrates the effectiveness of the recognition-oriented aspect of our face hallucination algorithm, even when each subject has only two samples in the training set.

E. Comparison with Deep Learning-Based Methods

We further evaluate our algorithm by comparing its performance on color images with several popular and/or successful deep learning-based approaches. We consider SRCNN [60] as a baseline algorithm along with DSCCN [61], DBPN [62], ESRGAN [63], SPSS [64], SRGAN [65], realESRGAN [66], SwinIR [67], and PULSE [35], of which the latter is a recently developed face hallucination technique. CNN-based methods were trained using our data, whereas the remaining networks
TABLE V

| Algorithm | PSNR | SSIM |
|-----------|------|------|
| SRCNN     | 20.52| 0.6866 |
| DCSCN     | 30.33| 0.8398 |
| DBPN      | 30.56| 0.8354 |
| ESRGAN    | 28.20| 0.7940 |
| SPSR      | 28.28| 0.7535 |
| BSRGAN    | 25.66| 0.6628 |
| realESRGAN| 24.44| 0.6719 |
| SwinIR    | 22.78| 0.5586 |
| Proposed  | 35.01| 0.9442 |

utilize their own pre-trained models. To alleviate this, the LR inputs provided for these networks are only the downsampled (and not blurred) version of the HR images. Our color results are obtained by performing the proposed method separately on each of the RGB channels. We perform the experiments on the Multi-PIE face database, and consider LR face images of size 16 × 12 with scaling factor 4. The PSNR and SSIM evaluation of the algorithms is reported in Table V and some hallucinated faces generated by different methods are also displayed in Fig. 17. It appears that SRCNN manages to enhance parts of the face images, but leaves undesired artifacts around the boundary regions. DCSCN and DBPN have done only slightly better than the bicubic interpolation, whereas the results produced by GAN-based algorithms are deformed and unclear. PULSE generates noise-free faces which hardly resemble their true identities. In general, despite achieving satisfactory results on higher resolutions, deep learning-based super-resolution algorithms seem to fail to enhance very low-resolution face images, with their results being vague and distorted.

F. Computational Complexity

The optimization procedure of the proposed algorithm consists of two phases. As discussed in section II, the \( l_2 - l_2 \) minimization problem can be solved through a closed-form solution, hence the \( l_1 \)-minimization problem is basically the most time-consuming part of the reconstruction process, which can also be solved efficiently using various \( l_1 \)-optimization approaches. Overall, the proposed method is considered to be a very fast algorithm, with not more than 30 iterations required for it to converge. To compare the computational time of our algorithm with those of the other methods and evaluate the effects of different parameters on its runtime, we perform experiments on the AR and the FERET databases with LR input size and scaling factor of 10 × 8 and 4, respectively, using MATLAB 2020b and a computer with 6GB memory and 1.8 GHz CPU. Table VI presents the average runtime of each method on the AR database when the training set size is 100. One can notice that the proposed algorithm performs face hallucination with reasonable computational cost compared to the competitive methods. We also measure the runtime of each algorithm with respect to the dataset size and the scaling factor. The results, which are displayed in Fig. 18 reveal that the computational time of our algorithm is not much affected by both parameters, whereas the ones for TLcR-RL \([22]\) and SSR \([57]\) grow exponentially, making them practically inefficient in the case of large images or big datasets. It should also be noted that reducing the number of iterations – which was previously shown to be slightly significant to the quality of the hallucinated face image after the first few iterations – will decrease the computational cost of the proposed algorithm even further. One can also think of applying collaborative representation \([68]\) instead of sparse representation, which leads to a super-fast face hallucination procedure with both subproblems having closed-form solutions.

G. Pose-Robust Face Hallucination

In order to evaluate the efficiency of the proposed 3D dictionary alignment procedure introduced in section II, extensive experiments were conducted on the Multi-PIE and the LFW face databases. Since the methods used in the frontal face experiments are unable to perform pose-robust face hallucination, we integrate them with our 3D dictionary alignment scheme, hence they use the same aligned dictionary to perform face hallucination. Moreover, the parameters and settings associated with all the algorithms remain the same as in the previous experiments. We use \([52]\) to extract 3D landmarks from both the LR and HR faces, and \([51]\) to perform 3D face reconstruction. To apply face alignment, all 68 landmark points were taken into consideration.

1) The Multi-PIE Dataset: For each of the subjects in the Multi-PIE face database, we consider illumination condition 10 and select the frontal face images (camera 05-1) taken in neutral position from all sessions as the training samples (thus, there are five samples per subject in the training set), and randomly select one sample from their images in the same settings but under different pose variations (cameras 04-1, 05-
Fig. 17. Qualitative comparison of the proposed method with various deep learning-based face hallucination and image super-resolution algorithms on the 16 × 12 test faces of the Multi-PIE database with scaling factor 4. (a) LR input. (b) Bicubic interpolation. (c) SRCNN [60]. (d) DCSCN [61]. (e) DBPN [62]. (f) ESRGAN [63]. (g) SPSR [64]. (h) BSRGAN [65]. (i) realESRGAN [66]. (j) SwinIR [67]. (k) PULSE [35]. (l) Proposed. (m) Ground truth.

Fig. 18. Computational time variations of different approaches according to (a) Scaling factor, and (b) Dataset size.

TABLE VII
| Algorithm | Multi-PIE | LFW |
|-----------|-----------|-----|
|           | PSNR      | SSIM| PSNR | SSIM |
| Bicubic   | 22.84     | 0.6466 | 21.65 | 0.6876 |
| Wang [6]  | 27.46     | 0.8038 | 25.77 | 0.8018 |
| LSR [18]  | 29.26     | 0.8522 | 27.66 | 0.8469 |
| LcR [20]  | 29.52     | 0.8534 | 27.57 | 0.8440 |
| LINE [21] | 29.83     | 0.8608 | 27.95 | 0.8543 |
| SSR [57]  | 29.25     | 0.8516 | 27.43 | 0.8413 |
| LM-CSS [15]| 28.91 | 0.8517 | 27.11 | 0.8395 |
| TRNR [58] | 30.23     | 0.8712 | 28.47 | 0.8661 |
| TLcR-RL [22]| 30.65 | 0.8843 | 28.88 | 0.8757 |
| Proposed  | **32.07** | **0.8950** | **31.01** | **0.9149** |

The performance of the proposed method is further evaluated in a real-world scenario, where samples in both the training and test sets are taken in uncontrolled conditions. Among the face images in the LFW database, we consider those belonging to the subjects with 10 to 14 samples in the database, which forms a subset containing 73 subjects and 894 samples. Since the images in the LFW dataset are taken in the wild, unlike the previous experiment, training samples may contain various degradations such as occlusion and illumination variations, which makes the problem even more challenging. Although the proposed alignment procedure is expected to perfectly handle the difficulties related to pose variations, in some infrequent cases, there is a considerable difference in poses of a training sample and the test face, making the registration unfavorable for the task of face hallucination. This
where $\text{hist}$ returns the histogram of the input image, and $I$ and $I'$ are original and transformed faces, respectively. We empirically consider $\theta = 100$. Also, the size of the LR faces is $20 \times 20$ (similar to the previous experiment, the actual face regions are considerably smaller), and the scaling factor is 4. The PSNR and SSIM scores obtained by different methods (Table VII) shows differences of 2.13 dB and 0.0392 in favor of the proposed algorithm, respectively. Also from Fig. 21, one can observe that the proposed method has added much more information to the reconstructed faces compared to the competitive algorithms. This not only once again highlights the superiority of the proposed face hallucination algorithm over the other approaches, but also shows how efficient our 3D dictionary alignment method is, even when both the training and testing images are taken in uncontrolled conditions.

IV. Conclusion

This paper presented a fast and robust face super-resolution algorithm which, different from most of the existing methods, attempts to recover true individual-specific attributes of the low-resolution input face. This is achieved by introducing a MAP estimator which encourages the reconstructed image to lie near to the subspace spanned by the training samples of the subject to which the input face belongs. Our experiments indicated that not more than three samples per subject are required to form such a subspace and obtain clear, detailed, and artifact-free outputs, even in datasets with high variations in facial expressions. We further extended our method by proposing a 3D dictionary alignment framework which proved to be vastly effective in real-world scenarios. Our evaluations revealed that, owing to its efficient closed-form based optimization procedure, the proposed algorithm is a very fast method which shows impressive robustness against face degradations. The identity-preserving aspect of our algorithm also demonstrated its significance in the task of low-resolution face recognition, even when there were only two samples of each subject available. The comparison results of the proposed method with some competitive algorithms, including both position-patch and deep learning-based methods, illustrated its superiority over the state-of-the-art face hallucination techniques in terms
of both quantitative measurements and visual impressions. One possible extension of our proposed algorithm is to employ more robust sparse representation-based regularization terms (e.g., [69] and [70]) to enhance its flexibility and robustness. Our future work will also focus on introducing the blurring operator to the main objective function as a variable and designing a blind face hallucination scheme.

REFERENCES

[1] K. Nasrollahi and T. B. Moeslund, “Super-resolution: a comprehensive survey,” Mach. Vis. Appl., vol. 25, pp. 1423–1468, 2014.

[2] W. W. W. Zou and P. C. Yuen, “Very low resolution face recognition problem,” IEEE TIP, vol. 21, no. 1, pp. 327–340, 2012.

[3] S. Baker and T. Kanade, “Hallucinating faces,” in Int. Conf. Autom. Face Gesture Recognit., 2000, pp. 83–88.

[4] C. Su, Y. Zhuang, L. Huang, and F. Wu, “Steerable pyramid-based face hallucination,” Pattern Recognit., vol. 38, no. 6, pp. 813–824, 2005.

[5] C. Liu, H. Y. Shum, and C. S. Zhang, “A two-step approach to hallucinating faces: Global parametric model and local nonparametric model,” in CVPR, vol. 1, 2001, pp. 192–198.

[6] X. Wang and X. Tang, “Hallucinating face by eigentransformation,” IEEE TSMCC, vol. 35, no. 3, pp. 425–434, 2005.

[7] Y. Zhuang, J. Zhang, and F. Wu, “Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation,” Pattern Recognit., vol. 40, no. 11, pp. 3178–3194, 2007.

[8] J. Yang, H. Tang, Y. Ma, and T. Huang, “Face hallucination via sparse coding,” in ICIP, 2008, pp. 1264–1267.

[9] J. S. Park and S. W. Lee, “An example-based face hallucination method for single-frame, low-resolution facial images,” IEEE TIP, vol. 17, no. 10, pp. 1806–1816, 2008.

[10] W. Liu, D. Lin, and X. Tang, “Hallucinating faces: TensorPatch super-resolution and coupled residue compensation,” in CVPR, vol. 2, 2005, pp. 478–484.

[11] W. Fan and D. Y. Yeung, “Image hallucination using neighbor embedding over visual primitive manifolds,” in CVPR, 2007, pp. 1–7.

[12] H. Huang, H. He, X. Fan, and J. Zhang, “Super-resolution of human face image using canonical correlation analysis,” Pattern Recognit., vol. 43, no. 7, pp. 2532–2543, 2010.
fusion network for face image hallucination,” *IEEE TNNLS*, pp. 1–14, 2020.

[28] X. Yu and F. Porikli, “Ultra-resolving face images by discriminative generative networks,” in *ECCV*, 2016, pp. 318–333.

[29] X. Yu and F. Porikli, “Face hallucination with tiny unaligned images by transformative discriminative neural networks,” in *AAAI*, vol. 31, no. 1, Feb 2017, pp. 1–16.

[30] X. Yu and F. Porikli, “Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders,” in *CVPR*, 2017, pp. 5367–5375.

[31] X. Yu, F. Porikli, B. Fernando, and R. Hartley, “Hallucinating unaligned face images by multiscale transformative discriminative networks,” in *Int. J. Comput. Vis.*, vol. 128, 2020, pp. 520–526.

[32] A. Bulat, J. Yang, and G. Tzimiropoulos, “To learn image superresolution, use a GAN to learn how to do image degradation first,” in *ECCV*, 2018, pp. 187–202.

[33] L. Yang et al., “HiFaceGAN: Face renovation via collaborative suppression and replenishment,” in *ACMMM*, 2020, pp. 1551–1560.

[34] C. Chen, D. Gong, H. Wang, Z. Li, and K. Y. K. Wong, “Learning spatial attention for face super-resolution,” *IEEE TIP*, vol. 30, pp. 1219–1231, 2021.

[35] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, “PULSE: Self-supervised photo upsampling via latent space exploration of generative models,” in *CVPR*, 2020, pp. 2223–2232.

[36] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, “ATMFN: Adaptive-threshold-based multi-modal fusion network for compressed face hallucination,” *IEEE TMM*, vol. 22, no. 10, pp. 2734–2747, 2020.

[37] L. Chen, J. Pan, and Q. Li, “Robust face image super-resolution via joint learning of subdivided contextual model,” *IEEE TIP*, vol. 28, no. 12, pp. 5897–5909, 2019.

[38] K. Su, Q. Tian, Q. Xue, N. Sebe, and J. Ma, “Neighborhood issue in single-frame image super-resolution,” in *ICME*, Jul 2005, pp. 4–8.

[39] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2001.

[40] Y. Romano and M. Elad, “Con-patch: When a patch meets its context,” *IEEE TIP*, vol. 25, no. 9, pp. 3967–3978, 2016.

[41] C. Hsu, C. W. Lin, W. T. Su, and G. Cheung, “SiGAN: Siamese generative adversarial network for identity-preserving face hallucination,” *IEEE TIP*, vol. 28, no. 12, pp. 6225–6236, 2019.

[42] M. Zhang and Q. Ling, “Supervised pixel-wise GAN for face super-resolution,” *IEEE TMM*, vol. 23, pp. 1938–1950, 2021.

[43] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE TPAMI*, vol. 30, no. 2, pp. 210–227, 2009.

[44] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J. Y. Tourneret, “Fast single image super-resolution using a new analytical solution for l1–l2 problems,” *IEEE TIP*, vol. 25, no. 8, pp. 3683–3697, 2016.

[45] A. Chakrabarti, A. N. Rajagopalan, and R. Chellappa, “Super-resolution of face images using kernel PCA-based prior,” *IEEE TMM*, vol. 9, no. 4, pp. 888–892, 2007.

[46] K. Jia and S. Gong, “Generalized face super-resolution,” *IEEE TIP*, vol. 17, no. 6, pp. 873–886, 2008.

[47] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, “Close the loop: Joint blind image restoration and recognition with sparse representation prior,” in *IEEE ICCV*, 2011, pp. 770–777.

[48] Q. Wei, N. Dobigeon, and J. Y. Tourneret, “Fast fusion of multi-band images based on solving a sylvester equation,” *IEEE TIP*, vol. 24, no. 11, pp. 4109–4121, 2015.

[49] W. W. Hager, “Updating the inverse of a matrix,” *SIAM Review*, vol. 31, no. 2, pp. 221–239, 1989.

[50] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[51] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, “Joint 3D face reconstruction and dense alignment with position map regression network,” in *ECCV*, 2018, pp. 534–551.

[52] J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Z. Li, “Towards fast, accurate and stable 3D dense face alignment,” in *ECCV*, Aug 2020, pp. 152–168.

[53] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *IEEE TPAMI*, vol. 22, no. 10, pp. 1090–1104, 2000.

[54] J. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PfE,” *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.

[55] A. Martinez and R. Benavente, “The AR face database: CVC technical report,” 24, 1998.

[56] G. B. Huang, M. Ramesh, T. Berg, and E. L. Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” in *Workshop on faces in ‘Real-Life’ images: detection, alignment, and recognition*, Oct. 2008.

[57] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, “Noise robust face image super-resolution through smooth sparse representation,” *IEEE TCBR*, vol. 47, no. 11, pp. 3991–4002, 2017.

[58] J. Jiang, C. Chen, K. Huang, Z. Cai, and R. Hu, “Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation,” *Inf. Sci.*, vol. 367–368, pp. 354–372, 2016.

[59] A. Martinez and A. C. Kak, “PCA versus LDA,” *IEEE TPAMI*, vol. 23, no. 2, pp. 228–233, 2001.

[60] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE TPAMI*, vol. 38, no. 2, pp. 295–307, 2016.

[61] J. Yamanaka, S. Kuwashima, and T. Kurita, “Fast and accurate image super resolution by deep CNN with skip connection and network in network in *ICONIP*, June 2017, pp. 217–225.

[62] M. Haris, G. Shakhnarovich, and N. Ukita, “Deep back-projection networks for super-resolution,” in *CVPR*, June 2018.

[63] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *ECCV*, September 2018.

[64] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, “Structure-preserving super resolution with gradient guidance,” in *CVPR*, 2020, pp. 7766–7775.

[65] K. Zhang, J. Liang, L. V. Gool, and R. Timofte, “Designing a practical degradation model for deep blind image super-resolution,” *ICCV*, 2021.

[66] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,” *ICS*, pp. 505–512, 2018.

[67] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, “SwinIR: Image restoration using swin transformer,” *ICCVW*, 2021.

[68] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in *ICCV*, 2011, pp. 471–478.

[69] J. Jiang and J. Lai, “Sparse and dense hybrid representation via dictionary decomposition for face recognition,” *IEEE TPAMI*, vol. 37, no. 5, pp. 1067–1079, 2015.

[70] Y. Xu, Z. Zhong, J. Yang, J. You, and D. Zhang, “A new discriminative sparse representation method for robust face recognition via l2 regularization,” *IEEE TNNLS*, vol. 28, no. 10, pp. 2233–2242, 2017.