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

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

20-11-2021 / 29-11-2021

CITATION

Abbasi, Ali; Rahmati, Mohammad (2021): Identity-Preserving Pose-Robust Face Hallucination Through Face Subspace Prior. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.17055734.v1>

DOI

[10.36227/techrxiv.17055734.v1](https://doi.org/10.36227/techrxiv.17055734.v1)

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 be 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 have 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. Yu *et al.* [24] pioneered in GAN-based face hallucination algorithms, by considering an architecture which consisted of two discriminative and generative networks. To avoid oversmoothing problem and preserve more textural details, WaveletSRNet [25] reconstructed HR images in wavelet coefficient domain. Chen *et al.* [26], extracted multi-scale features by incorporating multiple encoders and decoders in bottom-up and top-bottom patterns. In [27], a super-resolution technique was suggested which decomposed faces and recovered different components. Jiang *et al.* [28] developed a network with two individual branches to learn global facial shape and local facial components, whereas [29] 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

Fig. 1. Effects of the patch size on the subject of the most similar training patches. Four different sized LR patches (denoted by different colors) are written as the linear combinations of the training patches in the same corresponding positions, whereas training images with green and red boundaries indicate identical and different subjects with reference to the test image subject, respectively. The top three most similar patches as well as the position of three patches belonging to the test image subject are displayed. When the whole image is taken into account, the top three most similar patches all belong to the subject of the LR test image, whereas the reverse is the case with smaller patches. Furthermore, considering the position of the training patches in each case (i.e., the coefficients), one can notice that the similarity between the test patch and those patches which belong to the test image subject is directly correlated with the patch size.

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 [30]. 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) [31] 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 [32], the size of the training set should grow exponentially with the patch size to guarantee valid matches [33] and avoid ghosting effects [22]. Although several approaches have been suggested to alleviate this problem [21], [22], [30], 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

Fig. 2. Influence of the patch size on the average neighborhood preservation rates between the LR and HR image manifolds, based on different levels of degradation (i.e., blur kernel size) and when the number of neighbors K is set to 200. Patch-based strategies tend to be more erroneous as the patches decrease in size, and this becomes even worse in the presence of higher levels of degradation. Considering the entire face image as a patch leads to a sharp increase in the level of NPR and more robustness against facial degradation.

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 [34]. 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 [35]. 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 [34]. In recent years, several network architectures and loss functions have been suggested to incorporate the identity prior into the learning procedure [34], [35], 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* [36]. 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 [37].

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

Fig. 3. The results obtained by a recently introduced GAN-based face hallucination technique [38] and the proposed algorithm in a real-world problem. Despite reconstructing aesthetically pleasing faces, the facial details recovered by deep learning methods may sometimes be vastly different from the ones in the ground truth face image. (a) Low-resolution input. (b) Bicubic interpolation. (c) PULSE [38]. (d) Proposed method. (e) Ground truth image.

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.

The organization of the rest of the paper is as follows: in section II, the face subspace prior along with our main MAP-based model are explained. 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.

II. PROPOSED METHOD

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:

$$\mathbf{y} = \mathbf{S}\mathbf{H}\mathbf{x} + \epsilon \quad (1)$$

in which $\mathbf{y} \in \mathbb{R}^{m_l \times 1}$ is the low-resolution input with $\mathbf{x} \in \mathbb{R}^{m_h \times 1}$ as its high-resolution equivalent, where $m_l = w_l \times h_l$ and $m_h = w_h \times h_h$. In addition, $\mathbf{H} \in \mathbb{R}^{m_h \times m_h}$ represents blurring filter, $\mathbf{S} \in \mathbb{R}^{m_l \times m_h}$ 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 $\mathbf{x}$ as below:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{S}\mathbf{H}\mathbf{x}\|_2^2 \quad (2)$$

which leads to the following solution:

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

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 $\mathbf{H}^T \mathbf{S}^T \mathbf{S} \mathbf{H}$ is singular, the problem is also ill-posed with infinite space of possible solutions. Moreover, solving (3) requires inverting the matrix $\mathbf{H}^T \mathbf{S}^T \mathbf{S} \mathbf{H}$ with a computational complexity of the order $\mathcal{O}(m_h^3)$, which makes it practically inefficient in many real scenarios [37].

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:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{S}\mathbf{H}\mathbf{x}\|_2^2 + \xi \Phi(\mathbf{R}\mathbf{x}) \quad (4)$$

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 ξ which determines the contribution of each term. $\mathbf{R}$ 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 [39] or patch-based [40] 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 [32]. 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 [36]. Let $\mathbf{D}_{h,i} = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,n_i}]$ 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 $\mathbf{x}$ belongs to the subject i , for some scalar coefficients $\alpha_{i,j} \in \mathbb{R}$, $j = 1, 2, \dots, n_i$, it can be represented as $\mathbf{x} = \alpha_{i,1}\mathbf{x}_{i,1} + \alpha_{i,2}\mathbf{x}_{i,2} + \dots + \alpha_{i,n_i}\mathbf{x}_{i,n_i}$. Therefore, in case the subject of the input LR face is given beforehand, a suitable prior term for (4) would be

$$\Phi(\mathbf{x}) = \|\mathbf{x} - \alpha_{i,1}\mathbf{x}_{i,1} - \alpha_{i,2}\mathbf{x}_{i,2} - \dots - \alpha_{i,n_i}\mathbf{x}_{i,n_i}\|_2^2 \quad (5)$$

To generalize the above term, a dictionary matrix $\mathbf{D}_h$ 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, $\mathbf{D}_h = [\mathbf{D}_{h,1}, \mathbf{D}_{h,2}, \dots, \mathbf{D}_{h,c}] = [\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, \dots, \mathbf{x}_{c,n_c}]$. Therefore, (5) changes to

$$\Phi(\mathbf{x}) = \|\mathbf{x} - \mathbf{D}_h \boldsymbol{\alpha}\|_2^2 \quad (6)$$

where $\boldsymbol{\alpha}$ 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

$$\{\hat{\mathbf{x}}, \hat{\boldsymbol{\alpha}}\} = \arg \min_{\mathbf{x}, \boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{S}\mathbf{H}\mathbf{x}\|_2^2 + \mu \|\mathbf{x} - \mathbf{D}_h \boldsymbol{\alpha}\|_2^2 \quad (7)$$

Provided that the number of subjects is sufficiently large, $\boldsymbol{\alpha}$ is expected to have a sparse representation, since only few of its elements will have nonzero values. Therefore, a new regularization term $\|\boldsymbol{\alpha}\|_0$ should be included in (7), and since this leads to an NP-hard problem, the l_0 -norm is replaced by l_1 -norm [36]:

$$\{\hat{\mathbf{x}}, \hat{\boldsymbol{\alpha}}\} = \arg \min_{\mathbf{x}, \boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{S}\mathbf{H}\mathbf{x}\|_2^2 + \mu \|\mathbf{x} - \mathbf{D}_h \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (8)$$

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

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

$$\mathbf{x}_{t+1} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{S}\mathbf{H}\mathbf{x}\|_2^2 + \mu \|\mathbf{x} - \mathbf{D}_h \boldsymbol{\alpha}_t\|_2^2 \quad (9)$$

whose closed-form solution is given by

$$\mathbf{x}_{t+1} = (\mathbf{H}^T \mathbf{S}^T \mathbf{S} \mathbf{H} + 2\mu \mathbf{I}_{m_h})^{-1} (\mathbf{H}^T \mathbf{S}^T \mathbf{y} + 2\mu \mathbf{D}_h \boldsymbol{\alpha}_t) \quad (10)$$

Unlike the optimization procedure of other similar inverse problems (e.g., image deblurring [41]) which can be solved efficiently in the frequency domain, here, the presence of the decimation operator $\mathbf{S}$ in the fidelity term, and the fact that the product matrix $\mathbf{SH}$ 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, [37] showed that under certain assumptions on the decimation operator $\mathbf{S}$ and the blurring matrix $\mathbf{H}$, the optimization problem admits a closed-form solution in the frequency domain.

More precisely, assuming $\mathbf{H}$ as the matrix representation of the cyclic convolution operator, one can decompose the blurring operator $\mathbf{H}$ and its conjugate transpose $\mathbf{H}^H$ as

$$\mathbf{H} = \mathbf{F}^H \mathbf{\Lambda} \mathbf{F}, \quad \mathbf{H}^H = \mathbf{F}^H \mathbf{\Lambda}^H \mathbf{F} \quad (11)$$

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

Additionally, $\mathbf{S}$ is assumed to be a downsampling operator whose conjugate transpose $\mathbf{S}^H$ interpolates the decimated image with zeros, and satisfies the relationship $\mathbf{S}\mathbf{S}^H = \mathbf{I}_{m_l}$. By considering $\mathbf{S} \triangleq \mathbf{S}^H\mathbf{S}$, which operates as an element-wise multiplication by an $m_h \times m_h$ matrix with ones at the sampled positions and zeros elsewhere, [42] showed that:

$$\mathbf{F}\mathbf{S}\mathbf{F}^H = \frac{1}{d^2} \mathbf{J}_{d^2} \otimes \mathbf{I}_{m_l} \quad (12)$$

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

$$\mathbf{x}_{t+1} = \mathbf{F}^H \left(\frac{1}{d^2} \mathbf{\Lambda}^H \mathbf{\Lambda} + 2\mu \mathbf{F}\mathbf{F}^H \right)^{-1} \mathbf{F}(\mathbf{H}^H \mathbf{S}^H \mathbf{y} + 2\mu \mathbf{D}_h \alpha_t) \quad (13)$$

where $\mathbf{\Lambda} \in \mathbb{C}^{m_l \times m_h}$ is defined as $\mathbf{\Lambda} = [\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \dots, \mathbf{\Lambda}_{d^2}]$, with the blocks $\mathbf{\Lambda}_i \in \mathbb{C}^{m_l \times m_l}$ ($i = 1, \dots, d^2$) such that $\text{diag}\{\mathbf{\Lambda}_1, \dots, \mathbf{\Lambda}_{d^2}\} = \mathbf{\Lambda}$. This can be further simplified by incorporating the Woodbury matrix identity [43] into (13) and obtaining the following closed-form solution:

$$\mathbf{x}_{t+1} = \frac{1}{2\mu} \left(\mathbf{r} - \mathbf{F}^H \mathbf{\Lambda}^H (2\mu d^2 \mathbf{I}_{m_l} + \mathbf{\Lambda} \mathbf{\Lambda}^H)^{-1} \mathbf{\Lambda} \mathbf{F} \mathbf{r} \right) \quad (14)$$

where $\mathbf{r} = \mathbf{H}^H \mathbf{S}^H \mathbf{y} + 2\mu \mathbf{D}_h \alpha_t$.

2) *Optimizing for α* : The coefficient vector α is updated using the following well-known l_1 -minimization problem:

$$\alpha_{t+1} = \arg \min_{\alpha} \|\mathbf{x}_t - \mathbf{D}_h \alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (15)$$

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

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

Algorithm 1 Identity-Preserving Face Hallucination

Input: HR and LR dictionaries $\mathbf{D}_h$ and $\mathbf{D}_l$, low-resolution image $\mathbf{y}$, blurring matrix $\mathbf{H}$, scaling factor d , regularization parameters μ and λ , number of iterations T .

- 1: Initialize α : $\alpha_0 = \arg \min_{\alpha} \|\mathbf{y} - \mathbf{D}_l \alpha\|_2^2 + \lambda \|\alpha\|_1$.
- 2: Factorize the blurring matrix: $\mathbf{H} = \mathbf{F}^H \mathbf{\Lambda} \mathbf{F}$.
- 3: Compute $\mathbf{\Lambda}$: $\mathbf{\Lambda} = [(\mathbf{1}_d^T \otimes \mathbf{I}_{w_l}) \otimes (\mathbf{1}_d^T \otimes \mathbf{I}_{h_l})] \mathbf{\Lambda}$.
- 4: **for** $t = 0 : T - 1$ **do**
- 5: Find the FFT of $\mathbf{r}$: $\mathbf{r} = \mathbf{F}(\mathbf{H}^H \mathbf{S}^H \mathbf{y} + 2\mu \mathbf{D}_h \alpha_t)$.
- 6: Find entrywise product in the frequency domain:
- 7: $\mathbf{x}_f = (\mathbf{\Lambda}^H (2\mu d^2 \mathbf{I}_{m_l} + \mathbf{\Lambda} \mathbf{\Lambda}^H)^{-1} \mathbf{\Lambda}) \mathbf{r}$.
- 8: Update $\mathbf{x}$: $\mathbf{x}_{t+1} = \frac{1}{2\mu} (\mathbf{r} - \mathbf{F}^H \mathbf{x}_f)$.
- 9: Update α : $\alpha_{t+1} = \arg \min_{\alpha} \|\mathbf{x}_t - \mathbf{D}_h \alpha\|_2^2 + \lambda \|\alpha\|_1$.
- 10: **end for**

Output: Hallucinated face $\hat{\mathbf{x}} = \mathbf{x}_{t+1}$.

C. 3D Dictionary Alignment

As face images of different poses are distributed on a highly nonlinear manifold [44], 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 [45] to generate 3D geometries associated with each of the training faces, and obtain two sets $\{M_i\}_{i=1}^n$ and $\{C_i\}_{i=1}^n$, such that $M_i = (V_i, T_i)$ is the 3D mesh associated with the i th training face, with $V_i \in \mathbb{R}^{n_v \times 3}$, $T_i \in \mathbb{N}^{n_t \times 3}$, and $C_i \in \mathbb{R}^{n_v \times 3}$ 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 [46] 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

Fig. 4. Flowchart of the 3D dictionary alignment framework. The 3D facial landmarks are first extracted from the upsampled version of the LR input face, before being used to align it with respect to a set of landmark points defined as reference and obtain the associated transformation matrix. Having previously calculated the transformations between the dictionary samples and the reference landmarks, the transformation matrices between the training samples and the input face image can therefore be efficiently obtained. After applying these transformation matrices to their corresponding training face objects and performing 3D rendering, the aligned HR and LR dictionaries are obtained which will be later used along with the masked LR input in the process of face hallucination.

Fig. 5. The average normalized mean error achieved by a state-of-the-art landmark detector [46] on 20 randomly selected faces from the Multi-PIE dataset with pose variations 04-1, 05-0, 13-0, and 14-0, degraded by average filter of different sizes with various downsampling factors from their 80×60 HR images. The results of landmark detection on a test face with different levels of degradation as well as its ground truth landmark points are also presented. The figure suggests that the detector robustness is more than enough to be used in our dictionary alignment pipeline.

and their ground truth points is fairly negligible, even when face images as small as 15×15 pixels with 5×5 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 face pose. To speed up the process, a set of landmark points p_{ref} are considered as the reference, and the transformation matrices $\{\tau_{p_1, p_{ref}}, \tau_{p_2, p_{ref}}, \dots, \tau_{p_n, p_{ref}}\}$ are calculated in the training phase, in which $\tau_{p_i, p_{ref}}$ 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 τ_{p_{ref}, p_y} between the LR input face and the reference landmarks, and obtain $\{\tau_{p_1, p_y}, \tau_{p_2, p_y}, \dots, \tau_{p_n, p_y}\}$ simply by $\tau_{p_i, p_y} = \tau_{p_i, p_{ref}} \cdot \tau_{p_{ref}, p_y}$.

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 $\mathbf{D}_h^a$ and $\mathbf{D}_l^a$. 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^m , which will be later used along with $\mathbf{D}_h^a$ and $\mathbf{D}_l^a$ 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 $n_v = 43,867$ and $n_t = 86,906$, 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.

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 selected¹ with their parameters tuned so that they produce their optimal results. The experiments on frontal face hallucination are performed on the CMU Multi-PIE [47] and the AR [48] public face datasets, whereas the pose-robust face super-resolution algorithm is tested on the CMU Multi-PIE and the LFW [49] databases.

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

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 $\{\tau_{p_1, p_{ref}}, \tau_{p_2, p_{ref}}, \dots, \tau_{p_n, p_{ref}}\}$, low-resolution image $\mathbf{y}$.

- 1: $\mathbf{y}_{bic} \leftarrow \text{upscale}(\mathbf{y})$
- 2: Extract 3D facial landmark points from $\mathbf{y}_{bic}$.
- 3: Align $\mathbf{y}_{bic}$ with respect to the reference landmarks p_{ref} and obtain the transformation matrix τ_{p_{ref}, p_y} .
- 4: **for** $i = 1 : n$ **do**
- 5: $\tau_{p_i, p_y} = \tau_{p_i, p_{ref}} \cdot \tau_{p_{ref}, p_y}$
- 6: $V_i^a = V_i \circ \tau_{p_i, p_y}$
- 7: $\mathbf{x}_i^a \leftarrow \text{render}(V_i^a, T_i, C_i)$
- 8: Add $\mathbf{x}_i^a$ to the HR dictionary $\mathbf{D}_h^a$.
- 9: **end for**
- 10: $\mathbf{D}_l^a \leftarrow \text{degrade}(\mathbf{D}_h^a)$
- 11: Obtain $\mathbf{y}^m$ by applying the mask extracted from the average of the LR aligned faces to $\mathbf{y}$.

Output: Aligned dictionaries $\mathbf{D}_h^a$ and $\mathbf{D}_l^a$, masked LR input $\mathbf{y}^m$.

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 4×4 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 μ and λ 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) *The Multi-PIE dataset:* We first 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 12×9 , and three different scaling factors 4, 8, and 16 are considered. The quantitative results of different methods can be seen in Table I. 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 first 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.

2) *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 [52], 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 I, 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 μ and λ . 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 μ and changing λ values over the range $[0, 10^4]$ with an interval of 500, as plotted in Fig. 7, one can notice that the best performance is achieved when λ is roughly set to 10^3 . We next fix the value of λ and chose 20 different values for μ from the range $[0, 20]$. The variations of PSNR and SSIM (Fig. 8) suggest that values closer to zero are more desirable for this parameter. When μ 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

Fig. 6. Face hallucination results obtained by various methods. The first three rows indicate the results on the Multi-PIE dataset (scaling factors 4, 8, and 16, respectively) and the last three rows are those conducted on the AR database (scaling factors 4, 8, and 16, respectively). (a) LR input. (b) Bicubic interpolation. (c) Wang [6]. (d) LSR [18]. (e) LcR [20]. (f) LINE [21]. (g) SSR [50]. (h) LM-CSS [15]. (i) TRNR [51]. (j) TLcR-RL [22]. (k) Proposed. (l) Ground truth.

Fig. 7. Influence of the parameter λ on the face hallucination performance.

Fig. 8. Objective results based on different values of μ .

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. 9 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 [51] 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. 10, 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

TABLE I
OBJECTIVE COMPARISON OF DIFFERENT ALGORITHMS ON THE MULTI-PIE AND THE AR FACE DATABASES BY DIFFERENT SCALING FACTORS

Algorithm	CMU Multi-PIE						AR					
	$12 \times 9 \mid \times 4$		$12 \times 9 \mid \times 8$		$12 \times 9 \mid \times 16$		$5 \times 4 \mid \times 4$		$5 \times 4 \mid \times 8$		$5 \times 4 \mid \times 16$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	21.20	0.5967	18.83	0.4920	17.69	0.5564	17.14	0.3444	14.92	0.2041	13.70	0.2674
Wang [6]	25.85	0.8338	25.43	0.7588	25.21	0.7429	22.42	0.8223	20.91	0.6834	20.27	0.6091
LSR [18]	26.88	0.8548	24.57	0.7102	22.72	0.6498	22.98	0.8407	20.37	0.6359	18.69	0.5127
LcR [20]	28.08	0.8841	25.43	0.7493	22.93	0.6602	23.62	0.8580	19.44	0.6052	18.98	0.5344
LINE [21]	28.97	0.9042	25.53	0.7585	23.02	0.6727	24.61	0.8877	19.58	0.6367	19.11	0.5664
SSR [50]	28.39	0.8963	26.52	0.7908	24.97	0.7118	23.79	0.8562	21.31	0.6876	20.17	0.5844
LM-CSS [15]	27.48	0.8748	26.46	0.7861	26.18	0.7867	22.19	0.8121	20.52	0.6503	19.83	0.5683
TRNR [51]	30.25	0.9278	26.01	0.7790	23.36	0.6860	25.34	0.9042	21.61	0.7274	19.45	0.5625
TLcR-RL [22]	29.88	0.9211	27.66	0.8324	26.78	0.7920	24.59	0.8859	21.86	0.7225	21.04	0.6455
Proposed	31.34	0.9445	29.12	0.8769	28.06	0.8228	26.52	0.9228	24.23	0.8509	22.54	0.7621

Fig. 9. Effect of the number of samples per subject on the performance of different algorithms.

Fig. 11. Improvements of PSNR and SSIM values in different iterations.

Fig. 10. Qualitative influence of the number of samples per subject on the hallucinated face. (a)–(f) Reconstructed images by the proposed method (bottom) and TLcR-RL [22] (top) with 1, 2, 3, 4, 7, and 13 samples per subject, respectively. (g) Ground truth.

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. 11, 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., $\hat{\mathbf{x}} \approx \mathbf{D}_h \alpha_0$. 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 12×9 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. 12. 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 I, 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.

Fig. 12. Impact of blur kernel size on the performance of different algorithms.

Fig. 13. Comparison of quantitative scores of the proposed algorithm in its initial and final iterations on each specific test sample.

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 $\mathbf{D}_h \alpha_0$. Fig. 13 displays the quantitative indicators for each specific test face. 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. 14, 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 8×6 , whereas the inputs in the AR subset are of size 10×8 , and the scaling factor in both experiments is 4. LR faces are obtained by applying a 7×7 Gaussian filter with $\sigma = 2$ to the HR

Fig. 14. Visual comparison of the initial and final hallucinated faces. (a) Initial reconstruction. (b) Final reconstruction. (c) Ground truth.

TABLE II
EVALUATION OF FACE RECOGNITION PERFORMANCE ACHIEVED BY DIFFERENT ALGORITHMS ON THE MULTI-PIE AND THE AR DATABASES

Algorithm	Multi-PIE	AR
Wang [6]	34.11	52.00
LSR [18]	81.40	65.00
LcR [20]	79.07	70.00
LINE [21]	82.95	68.00
SSR [50]	65.89	73.00
LM-CSS [15]	66.67	57.00
TRNR [51]	81.40	73.00
TLcR-RL [22]	83.72	70.00
Proposed	92.25	76.00
HR	98.45	88.00

images, before downsampling them to the desired sizes. The resultant images of all methods are classified using the SRC classifier [36]. As for the proposed method, the final coefficient vector $\hat{\alpha}$ is used in the classification. Table II represents the recognition accuracy achieved by different approaches in both experiments, whereas Fig. 15 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. 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

Fig. 15. Cumulative recognition rates versus rank achieved by different methods. (a) The Multi-PIE dataset. (b) The AR dataset.

Fig. 16. Computational time variations of different approaches according to (a) Scaling factor, and (b) Dataset size.

algorithm with those of the other methods and evaluate the effects of different parameters on its runtime, we perform several experiments 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 III presents the average runtime of each method 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. 16, reveal that the computational time of our algorithm is not much affected by both parameters, whereas the ones for TLcR-RL [22] and SSR [50] 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 [53] instead of sparse representation, which leads to a super-fast face hallucination procedure with both subproblems having closed-form solutions.

TABLE III
COMPARISON OF THE AVERAGE RUNNING TIME OF DIFFERENT METHODS

Algorithm	Wang	LSR	LcR	LINE	SSR
Runtime (s)	0.11	0.13	1.04	0.93	10.03
Algorithm	LM-CSS	TRNR	TLcR-RL	Proposed	
Runtime (s)	1.07	1.02	11.48	1.05	

TABLE IV
POSE-ROBUST FACE HALLUCINATION PERFORMANCE ACHIEVED BY DIFFERENT ALGORITHMS ON THE MULTI-PIE AND THE LFW DATASETS

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 [50]	29.25	0.8516	27.43	0.8413
LM-CSS [15]	28.91	0.8517	27.11	0.8395
TRNR [51]	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

F. 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 [46] to extract 3D landmarks from both the LR and HR faces, and [45] 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, 05-0, 13-0, and 14-0) from session one to form the test set. Original images are cropped in such a way that they contain the entire face to facilitate the procedure of 3D face reconstruction (i.e., the face regions are usually much smaller than the actual image size). The LR face images are of size 20×20 and the scaling factor is 4. Table IV summarizes the results obtained by each method. Our algorithm outperforms the second best method by 1.42 dB in PSNR and 0.0107 in SSIM. The results shown in Fig. 17 also verify the effectiveness of our proposed dictionary alignment technique, which has clearly enhanced the performance of all the face super-resolution algorithms. Once again, by investigating the hallucinated faces, one can easily notice more recovered facial details in our results than those of the other approaches.

2) *The LFW Dataset (Face Hallucination in the Wild)*: 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

Fig. 17. Face hallucination results generated by different methods using the proposed 3D dictionary alignment scheme on the test samples of the Multi-PIE (first three rows) and the LFW (second three rows) face databases with scaling factor 4. (a) LR input. (b) Bicubic interpolation. (c) Wang [6]. (d) LSR [18]. (e) LcR [20]. (f) LINE [21]. (g) SSR [50]. (h) LM-CSS [15]. (i) TRNR [51]. (j) TLcR-RL [22]. (k) Proposed. (l) Ground truth.

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. The size of the LR faces is 20×20 (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 IV) shows differences of 2.13 dB and 0.0392 in favor of the proposed algorithm, respectively. Also from Fig. 17, 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 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., [54] and [55]) 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, and C. S. Zhang, "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.
- [13] Y. Hu, K. M. Lam, G. Qiu, and T. Shen, "From local pixel structure to global image super-resolution: A new face hallucination framework," *IEEE TIP*, vol. 20, no. 2, pp. 433–445, 2011.
- [14] B. Li, H. Chang, S. Shan, and X. Chen, "Aligning coupled manifolds for face hallucination," *IEEE SPL*, vol. 16, no. 11, pp. 957–960, 2009.
- [15] R. A. Farrugia and C. Guillemot, "Face hallucination using linear models of coupled sparse support," *IEEE TIP*, vol. 26, no. 9, pp. 4562–4577, 2017.
- [16] J. Shi, X. Liu, Y. Zong, C. Qi, and G. Zhao, "Hallucinating face image by regularization models in high-resolution feature space," *IEEE TIP*, vol. 27, no. 6, pp. 2980–2995, 2018.
- [17] J. Shi and G. Zhao, "Face hallucination via coarse-to-fine recursive kernel regression structure," *IEEE TMM*, vol. 21, no. 9, pp. 2223–2236, 2019.
- [18] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognit.*, vol. 43, no. 6, pp. 2224–2236, 2010.
- [19] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *IEEE SPL*, vol. 18, no. 6, pp. 367–370, 2011.
- [20] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE TMM*, vol. 16, no. 5, pp. 1268–1281, 2014.
- [21] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multi-layer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE TIP*, vol. 23, no. 10, pp. 4220–4231, 2014.
- [22] J. Jiang, Y. Yu, S. Tang, J. Ma, A. Aizawa, and K. Aizawa, "Context-patch face hallucination based on thresholding locality-constrained representation and reproducing learning," *IEEE TCYB*, vol. 50, no. 1, pp. 324–337, 2020.
- [23] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild," in *AAAI*, 2015.
- [24] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *ECCV*, 2016, pp. 318–333.
- [25] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution," in *ICCV*, Oct 2017, pp. 1689–1697.
- [26] Z. Chen, J. Lin, T. Zhou, and F. Wu, "Sequential gating ensemble network for noise robust multiscale face restoration," *IEEE TCYB*, vol. 51, no. 1, pp. 451–461, 2021.
- [27] X. Hu, P. Ma, Z. Mai, S. Peng, Z. Yang, and L. Wang, "Face hallucination from low quality images using definition-scalable inference," *Pattern Recognit.*, vol. 94, pp. 110–121, 2019.
- [28] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, "Dual-path deep fusion network for face image hallucination," *IEEE TNNLS*, pp. 1–14, 2020.
- [29] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, "ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination," *IEEE TMM*, vol. 22, no. 10, pp. 2734–2747, 2020.
- [30] 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.
- [31] 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.
- [32] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2001.
- [33] Y. Romano and M. Elad, "Con-patch: When a patch meets its context," *IEEE TIP*, vol. 25, no. 9, pp. 3967–3978, 2016.
- [34] C. 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.
- [35] M. Zhang and Q. Ling, "Supervised pixel-wise GAN for face super-resolution," *IEEE TMM*, vol. 23, pp. 1938–1950, 2021.
- [36] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE TPAMI*, vol. 31, no. 2, pp. 210–227, 2009.
- [37] 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 l_2 - l_2 problems," *IEEE TIP*, vol. 25, no. 8, pp. 3683–3697, 2016.
- [38] 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.
- [39] 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.
- [40] K. Jia and S. Gong, "Generalized face super-resolution," *IEEE TIP*, vol. 17, no. 6, pp. 873–886, 2008.
- [41] 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 *ICCV*, 2011, pp. 770–777.
- [42] 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.
- [43] W. W. Hager, "Updating the inverse of a matrix," *SIAM Review*, vol. 31, no. 2, pp. 221–239, 1989.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
- [48] A. Martinez and R. Benavente, "The AR face database: CVC technical report, 24," 1998.
- [49] 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.
- [50] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE TCYB*, vol. 47, no. 11, pp. 3991–4002, 2017.
- [51] 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.
- [52] A. Martinez and A. C. Kak, "PCA versus LDA," *IEEE TPAMI*, vol. 23, no. 2, pp. 228–233, 2001.
- [53] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *ICCV*, 2011, pp. 471–478.
- [54] X. 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.
- [55] Y. Xu, Z. Zhong, J. Yang, J. You, and D. Zhang, "A new discriminative sparse representation method for robust face recognition via l_2 regularization," *IEEE TNNLS*, vol. 28, no. 10, pp. 2233–2242, 2017.