Hallucinating Face Image by Regularization Models in High-Resolution Feature Space

Jingang Shi, Xin Liu, Yuan Zong, Chun Qi Member, IEEE and Guoying Zhao Senior Member, IEEE

Abstract—In this paper, we propose two novel regularization models in patch-wise and pixel-wise respectively, which are efficient to reconstruct high-resolution (HR) face image from low-resolution (LR) input. Unlike the conventional patch-based models which depend on the assumption of local geometry consistency in LR and HR spaces, the proposed method directly regularizes the relationship between the target patch and corresponding training set in the HR space. It avoids to deal with the tough problem of preserving local geometry in various resolutions. Taking advantage of kernel function in efficiently describing intrinsic features, we further conduct the patch-based reconstruction model in the high-dimensional kernel space for capturing nonlinear characteristics. Meanwhile, a pixel-based model is proposed to regularize the relationship of pixels in the local neighborhood, which can be employed to enhance the fuzzy details in the target HR face image. It privileges the reconstruction of pixels along the dominant orientation of structure, which is useful for preserving high-frequency information on complex edges. Finally, we combine the two reconstruction models into a unified framework. The output HR face image can be finally optimized by performing an iterative procedure. Experimental results demonstrate that the proposed face hallucination method produces superior performance than the state-of-the-art methods.

Index Terms—Face hallucination, super-resolution, regularization framework, manifold learning, kernel method.

I. INTRODUCTION

DURING the past decades, face analysis has received great interests and achieved impressive successes, especially in the applications under real-world environments. However, in many real-world applications, it often suffers from low-resolution problems because of the limitations of digital imaging devices, which can severely affect the performance of most face analysis algorithms. It is desirable to break the limitations of low-quality surveillance camera systems and further enhance the lost high-frequency details for practical applications.

Thus, we intend to approximate HR images from LR ones, using the super-resolution (SR) technique [1]. Generally, these algorithms can be mainly divided into two categories: interpolation-based methods [2]–[4] and learning-based methods [5]–[37]. The interpolation-based methods usually regularize the reconstruction process by using certain prior knowledge, while the learning-based methods aim to reconstruct the relationship between LR and HR images through a set of training pairs. For the problem of large magnification factor, the learning-based methods significantly improve the performance when they are compared with the interpolation-based ones. Due to the above advantages, learning-based methods attract more attention in recent research for practical applications. Chang et al. [10] utilized manifold learning methods to approximate the relationship of local patches. Yang et al. [16] introduced sparse coding methods to learn dual compact dictionaries, which are effective to represent and reconstruct HR patches. Timofte et al. [38], [39] employed regression-based methods for transforming the input LR patches into HR ones. A set of mapping functions are learned from training set to describe the reconstruction procedure. Dong et al. [5]–[7] exploited internal similarities of natural images and utilized repetitive patterns to refine HR images. Recently, deep learning methods [40]–[42] are also developed to generate HR images from LR ones in an end-to-end manner. The learning-based methods can be formulated for natural images, or can be designed to deal with particular tasks, such as face hallucination, medical image super-resolution and document image super-resolution. We focus on the learning-based face hallucination techniques in this paper.

According to the special characteristics of face images, Baker and Kanade [9] first proposed a learning-based face SR approach for reconstructing HR face images, which learns a prior on the distribution of image gradients through the training set. The specific SR procedure of face images is also called “face hallucination” in their literature. Liu et al. [12] performed the face hallucination work into two steps: global reconstruction and residual compensation. The first step generates the global face image which keeps the characteristics of face image using probabilistic method. Then, the second step produces the residual image in patch-wise to compensate the result of the first step. Following [12], lots of classical data models are utilized to describe the relationship of LR and HR face images, which includes PCA [11], Locality Preserving Projection (LPP) [13], Tensor Factorization [14], Non-negative Matrix Factorization (NMF) [16], Canonical Correlation Analysis (CCA) [15] and Two-dimensional Canonical Correlation Analysis (2DCCA) [17]. However, the reconstruction quality...
of global face images highly depends on the given training set in these methods. It requires a large number of training images for better results. Although residual compensation is utilized to alleviate the artifacts, ringing effects often take place in the high-frequency area of face images.

Recognizing the fact that the global reconstruction tends to induce artificial results, Ma et al. [19] proposed a patch-based reconstruction method instead of the complicated two-step model. In their method, face image is considered as a highly structured object and divided into patches according to the position prior. The position information can be served as a constraint, while the HR image patches are reconstructed by the combination of training patches at the same position. With the assumption that image patches share the same combination relationship in the LR and HR space, the combination weights can be learned in the LR space by a constrained least square problem. Similar idea was also further developed in [31] for the multiview hallucination problem. The additional position constraint really enhances the performance of face hallucination, which has been utilized by lots of researchers [20]–[30]. Some following methods devised reasonable regularization terms for achieving more stable and accurate solutions, such as $\ell_1$-norm minimization term [20] [24], constrained $\ell_p$-norm minimization term [25], fused Lasso-based smooth term [30]. Some researchers reconstructed the target HR image by transforming the original face image into latent subspace [18] or Discrete Cosine Transform (DCT) domain [22]. Manifold learning methods were further considered in [26] [27] to preserve the local geometry of image patches. The authors of [28] and [32] learned linear projection models from the training set in order to estimate the details of HR images. Zeng and Huang [33] expanded the training data for improving the quality of face hallucination results. Recently, deep learning has also been adopted in face hallucination problem [34]–[37]. Due to the advantages of deep learning, these methods produce fine performance with rich details. Meanwhile, more computational resources are also required in the training phase of deep neural network.

Generally, the common idea of patch-based methods is to represent the input LR patch as the combination of LR training patches with suitable reconstruction model. The reconstruction weights are then directly utilized to recover the target HR patch with the corresponding HR training set. An illustration is shown in Fig. 1. To preserve the consistency of reconstruction weights, these methods require a basic assumption that LR and HR patches share the same local geometry. Actually, due to the one-to-many relationship between LR patch and corresponding HR one, the combination weights in the LR space can hardly reflect the truth of HR space. This problem has also been presented in the literature [1], [18], [26], [27]. Previously, researchers attempted to improve the consistency of combination weights for achieving reasonable results. Though these algorithms devise further regularization to optimize the mismatch of local geometry, the correspondence between LR and HR space is really hard to completely satisfy. The weights obtained in the LR space can be considered as a sub-optimal solution, which gives a rough approximation to the original value. It is no doubt that plausible HR images can be reconstructed by directly utilizing the corresponding weights in the LR space. However, artifacts will take place in the target HR image since it fails to achieve the assumption of sharing the same reconstruction weights.

To improve the above disadvantages, we abandon the assumption of sharing the same weights in the space of various dimensions. In the proposed method, the relationship between the target patch and training images is only considered in the HR space, which avoids to deal with the tough problem of preserving local geometry in different resolutions. Furthermore, most previous algorithms utilize the reconstruction framework which represents each local patch as a linear combination of training patches. Nevertheless, it is a limitation to describe the reconstruction process by linear relationship when the number of training patches is large. Consequently, it is a good choice to leverage the nonlinearity for better representing the target patch. To this end, we embed the HR patches into an infinite-dimensional reproducing kernel Hilbert space (RKHS) by kernel function [43]. The HR patches are assumed to have a linear relationship in the RKHS, which implies the nonlinear relationship in the original HR space. Fig. 2 gives a concise instruction for the proposed regularization model, which is different from the traditional model shown in Fig. 1.

In the patch-based model, the whole HR image is usually divided into overlapped patches with moderate size (e.g.,
12 × 12). However, the patch size is a bit large for the reconstruction of local edge structure. The pixels along the edge are just a mere portion in such a patch, while the remaining pixels mainly describe the smooth texture. Since all the pixels in the patch are taken into account equally, it will induce somewhat blurring effect on the local structure. In other words, the pixels around the edge require further consideration to enhance the high-frequency information. Thus, we further propose an additional regularization in pixel-wise to constrain the distribution of pixels. For each pixel in the HR image, we regularize the relationship of neighboring pixels to describe local structure, while the edge details are enhanced by similar structures in the measured samples. The steering kernel [44] is employed to estimate the orientation of local structure in the neighborhood. The pixels on the dominant orientation will be further assigned with large weights in the pixel-based model, which is useful for suppressing artifacts on the complex edges. Finally, the above reconstruction models are incorporated into a global regularization framework, which produces the target HR image by iterative optimization. A flowchart of the proposed method is shown in Fig. 3. The contributions of this paper are summarized as follows:

- Due to the difficulty of preserving local geometry consistency, the assumption of sharing the same weights in various dimensional spaces is abandoned in the proposed method. Alternately, the patch-based model is directly built in the HR space, regardless of the combination relationship in the LR space.
- Motivated by the fact that kernel function can capture the nonlinearity of features, the HR patches are mapped into RKHS, which helps in efficiently representing the nonlinear relationship for the reconstruction.
- A pixel-based model is presented to emphasize influence on the pixels along the dominant orientation of structure, which results in preservation of details for the region of local edge.
- Experimental results show that the proposed method has the ability to produce fine performance with various experimental conditions. The results also illustrate the robustness of the proposed face hallucination method.

II. RELATED WORK

The target of face image hallucination is to estimate a HR face image from the corresponding LR one, where the low quality face image is degraded by smoothing, down-sampling and noising processes. The whole low-resolution face image generation process can be represented as:

\[ Y = DBX + \varepsilon \]  \hspace{1cm} (1)

where \( Y \) and \( X \) are the LR and HR face images respectively, \( B \) represents the blurring filter, \( D \) represents the down-sampling process and \( \varepsilon \) denotes the additional noise. The objective of reconstruction is to solve the above inverse problem in order to obtain the original HR face image \( X \).

Recently, plenty of algorithms [19]–[31] recover the HR face image from the corresponding LR one under the patch-based reconstruction framework due to its effectiveness. Generally, the patch-based methods first divide the whole LR image \( Y \) into overlapped patches. Then, the corresponding HR patches can be reconstructed by utilizing the LR and HR training pairs. The final output image is obtained by averaging all the hallucinated HR patches. For each LR patch \( y_i \), it will be represented by a linear combination of LR training samples \( \{y_i\}_{i=1}^P \):

\[ y = y_1\omega_1 + y_2\omega_2 + \ldots + y_p\omega_p \] \hspace{1cm} (2)

where \( \omega_i \) is the combination coefficient for the corresponding training sample. The HR patch \( x \) can be estimated by substituting the LR training patches \( \{y_i\}_{i=1}^P \) to the HR ones \( \{x_i\}_{i=1}^P \):

\[ x = x_1\omega_1 + x_2\omega_2 + \ldots + x_p\omega_p \] \hspace{1cm} (3)

However, there are two main disadvantages for these methods:

- The above reconstruction model assumes that image patches maintain similar local geometry in the LR and HR space. Thus, the LR and corresponding HR patches share the same linear combination coefficients for reconstruction. Nevertheless, the LR and HR patches actually fail to maintain the manifold consistency assumption. Define the function \( \varphi(x) \)
to describe the degradation procedure from the HR patch to the LR one:

\[ y_i = \varphi_i(x_i) \]  

Equation (2) can be rewritten as:

\[ \varphi(x) = \varphi_1(x_1)\omega_1 + \varphi_2(x_2)\omega_2 + \ldots + \varphi_p(x_p)\omega_p \]  

Though the degradation of the whole face image can be represented as a linear process according to (1), the generation of LR patch from the HR one is actually nonlinear due to the large receptive field of each LR pixel. An illustration is shown in Fig. 4. For most pixels in the LR patch, the range of receptive field exceeds the corresponding HR patch so that the function \( \varphi_1(\cdot) \) depends on some additional pixels out of the target area. What is worse, the function \( \varphi_i(\cdot) \) always varies for different patches \( x_i \) because of the influence of additional pixels in the receptive field. Apparently, it is hard to derive (3) from (5) when the expression of \( \omega \) is not explicit.

If we loose the strict condition and consider the function \( \varphi_i(\cdot) \) as a simple linear operator, the degradation of image patches can be represented as \( y_i = Px_i \) (\( P \in \mathbb{R}^{m \times n}, m < n \)). Thus, equation (2) can be rewritten as:

\[ Px = P_1\omega_1 + P_2\omega_2 + \ldots + P_p\omega_p \]  

If we want to derive (3) from (6), it requires the matrix \( P \in \mathbb{R}^{m \times n} \) to meet the condition \( \text{rank}(P) = n \). However, the number of rows is less than the number of columns in the matrix \( P \), which induces \( \text{rank}(P) \leq m < n \). Thus, it is not reasonable to reconstruct the HR patch with the manifold consistency assumption. In the proposed method, we will not consider the relationship of image patches in the LR space, but regularize the target patch in the HR space directly.

- In (3), it is assumed that each image patch can be reconstructed by a linear combination of patches in the training set. However, the linear model is difficult to describe the relationship between the target patch and the corresponding training patches. For the patch-based methods, face images are often divided into very small patches, such as only 3 × 3 pixels in LR case. The dimension of image patches is much less than the number of training samples, so that the solution of combination coefficients \( \{\omega_i\}_{i=1}^P \) is infinite. The previous approaches usually impose various regularization terms to obtain the unique solution. In this case, the results of \( \{\omega_i\}_{i=1}^P \) are highly influenced by the regularization prior instead of the original feature in the image patch. Thus, besides of utilizing the regularization prior, we also project the HR patches into an infinite dimensional RKHS for further reconstruction. Due to the increase of feature dimension, the optimization of combination relationship becomes more stable. Meanwhile, the nonlinear relationship of HR image patches is also captured in the RKHS, which produces better description for image feature than the linear model in the original HR space.

### III. The Proposed Algorithm

The reconstruction model (1) of the underlying HR face image \( X \) is severely under-determined, which means that each LR face image corresponds to infinite HR estimations. In order to improve this problem, it is essential to incorporate further regularization prior for the whole reconstruction procedure. With the additional prior, it is possible to obtain a unique solution for the HR face image \( X \). The reconstruction model can be represented as:

\[ X = \arg \min_X \left\{ \|Y - Dx\|^2 + \gamma R(X) \right\} \]  

where \( R(X) \) represents the regularization models and \( \gamma \) is the corresponding parameter that balances the reconstruction error and the regularization term. In (7), the first term ensures the consistency of the estimated HR face image and the observed LR one, while the additional prior further regularizes the local details for the HR face image. According to the characteristic of face image, we propose two effective regularization models in order to constrain the reconstruction process. The details of the above two priors will be described in the following.

#### A. Regularizing local texture by HR image patch

Let \( I^L_p \) and \( I^H_p \) represent the LR and HR training images respectively, where \( n = 1, 2 \ldots N \) is the corresponding sample number in the training set. As shown in Fig. 2, each face image is divided into overlapped patches according to the position. Denote \( L^p \) and \( H^p \) as the LR and HR image patches centered at position \( p \). The whole training set can be classified into \( M \) groups \( \{L_p, H_p\}_{p=1}^M \), where \( L_p = \{l^1_p, \ldots, l^N_p\} \) and \( H_p = \{h^1_p, \ldots, h^N_p\} \) represent the corresponding LR and HR patches located at the same position. With the position constraint, training pairs \( \{L_p, H_p\} \) can be utilized to recover the local details of LR patch at the corresponding position \( p \), which has been proved the effectiveness in [19].

For an input LR face \( Y \), it is also divided into overlapped patches for reconstruction. We represent the LR image patch centered at position \( p \) as \( y^L_p \), while the corresponding HR patches \( y^H_p \) is expected to be estimated. In the previous algorithms, the common method for reconstructing the HR parts is to assume that the HR and LR patches share the same geometry structure. The optimization problem can be solved by minimizing the following function:

\[ J(\omega_p) = \left\| y^L_p - \sum_i \omega^L_i l^i_p \right\|^2 + \lambda \Omega(\omega_p) \]  

where \( \omega_p \) is the vector that contains all the combination coefficients \( \omega_p^L \) and \( \lambda \) is utilized to balance the two data terms. The first term represents the reconstruction error in the LR space, and the second term enforces the constraint for the combination relationship.

In order to avoid the disadvantage for the inconsistency of LR and HR manifold, we directly consider the relationship between the target patch and training set in the HR space. The optimization objective can be written as:

\[ J(\omega_p, x^H_p) = \left\| x^H_p - \sum_i \omega^H_i h^i_p \right\|^2 + \lambda \Omega(\omega_p) \]  

In equation (9), the reconstruction weight is no longer estimated in the LR space, which avoids the inaccuracy of manifold consistency assumption.
In order to overcome the shortcoming of the above linear combination model between HR image patches, we further introduce kernel function to induce an infinite dimensional RKHS for capturing the nonlinear similarity of patches. All the image patches are projected into infinite dimensional kernel-mapped space, while the target is to find an optimal linear combination of nonlinear features. The linear relationship in such a RKHS leads to nonlinearity relationship for the HR patches in the original space. Since the nonlinear structures of patches are taken into account, it is more effective to estimate reasonable HR patches.

Let \( \phi : \mathbb{R}^d \rightarrow \mathbb{R}^F \) be a nonlinear mapping operator, which projects the data from the original feature space to an infinite dimensional RKHS space. The kernel function \( k(x_i, x_j) = (\phi(x_i))^T \phi(x_j) \) describes the nonlinear similarity between two features \( x_i \) and \( x_j \) in the RKHS. By using the operator \( \phi \), we are able to project the feature data from the original HR space into RKHS, where the target HR patch and the corresponding HR training patches can be represented as:

\[
x_p^H \rightarrow \phi(x_p^H), \quad H_p \rightarrow \phi(H_p) = \left[ \phi(h_p^1), \ldots, \phi(h_p^N) \right]
\]  

Thus, we substitute the kernel-mapped features to the formulation of objective function (9). It can be rewritten as:

\[
J(\omega_p, x_p^H) = \| \phi(x_p^H) - \sum_i \omega_p^i \phi(h_p^i) \|^2_2 + \lambda \Omega(\omega_p)
\]  

To avoid the bias solution, the regularization term \( \Omega(\omega_p) \) is required to constrain the combination in the RKHS. Inspired by [45], it will obtain better performance if local similar samples are privileged to get larger combination coefficients in the reconstruction. The similarity constraint is imposed on the RKHS, which can efficiently reflect the real characteristics of HR patches. Thus, we define a kernel-based similarity constraint which measures the distance of HR patches in the RKHS. The objective function is formulated as:

\[
J(\omega_p, x_p^H) = \| \phi(x_p^H) - \sum_i \omega_p^i \phi(h_p^i) \|^2_2 + \lambda \| d_p \odot \omega_p \|^2_2
\]  

s.t. \( \sum_i \omega_p^i = 1 \)  

(12)

where \( \odot \) denotes the Hadamard product, the constraint ensures the shift-invariant requirement and the vector \( d_p = [d_p^1, \ldots, d_p^N]^T \) serves as a penalty which describes the kernel similarities between the target HR patch and corresponding training samples in the kernel-mapped space. The kernel similarity distance can be represented as:

\[
d_p^i = \| \phi(x_p^H) - \phi(h_p^i) \|^2_2 = \sqrt{(\phi(x_p^H))^T \phi(x_p^H) + (\phi(h_p^i))^T \phi(h_p^i) - 2(\phi(x_p^H))^T \phi(h_p^i)}
\]  

\[
= \sqrt{k(x_p^H, x_p^H) + k(h_p^i, h_p^i) - 2k(x_p^H, h_p^i)}
\]  

Thus, if one training sample \( \phi(h_p^i) \) is similar with the target HR patch \( \phi(x_p^H) \), the penalty \( d_p^i \) would be small, which induces larger value for the corresponding weight \( \omega_p^i \). The above regularization term explicitly encourages the locality property for the reconstruction of patch-based model in the kernel-mapped space. Thus, there are only a few significant values for the solution, while most of the weight coefficients are close to zero. In practice, we merely consider the corresponding weights of \( P = 200 \) nearest neighboring samples, while set the coefficients of other samples to be zero. When the number of training samples is less than \( P \), all the samples are considered in the regularization.

Define \( R_p \) as the matrix to extract HR image patch centered at position \( p \) of the entire image. We then consider the whole HR image \( X \) as a variable. It leads to the following optimization problem:

\[
J(\omega_p, X) = \| \phi(R_p X) - \phi(H_p) \omega_p \|^2_2 + \lambda \| d_p \odot \omega_p \|^2_2
\]  

s.t. \( 1^T \omega_p = 1 \)  

(14)

where \( 1 \) is a column vector of ones. Incorporating (14) to the reconstruction model (7), the resulting optimization problem is represented as:

\[
\{\omega_p, X\} = \arg \min_{\omega_p, X} \{ \| Y - DBX \|^2_2
\]

\[+ \alpha \sum_p (\| \phi(R_p X) - \phi(H_p) \omega_p \|^2_2 + \lambda \| d_p \odot \omega_p \|^2_2) \}
\]  

s.t. \( 1^T \omega_p = 1, \ p = 1, 2, \ldots, M \)  

(15)

where the parameter \( \alpha \) measures the contribution of the corresponding regularization term.

**B. Regularizing local structure along edge**

The patch-based model recovers most high-frequency textures for the face image. However, the local structure requires further consideration for suppressing artifacts on the complex edges. It aims to characterize the relation between neighboring pixels for enhancing the edge structure, which is not considered in the patch-based prior.

For the pixel \( x_p \) on the \( p^{th} \) position of the estimated HR face image \( X \), a \( 5 \times 5 \) square window \( N_p \) is selected around the target pixel as the neighborhood. In order to describe the property of local structure, we search \( K = 50 \) nearest correlative neighbors from the same position in the HR training images for reference. The selected measured patches share similar structure for the texture near the edge, which can be utilized to optimize the local details for the target patch. Let \( t_p^i \) be the pixel at position \( p \) of the \( i^{th} \) correlative neighbor. The target pixel \( x_p \) is assumed to be estimated by linear regression:

\[
x_p = \sum_i t_p^i g_p^i
\]  

(16)

To obtain a unique solution, the locality-constrained \( \ell_2 \)-norm is utilized to regularize the regression:

\[
\beta_p = \arg \min_{\beta_p} \left[ \| x_p - \sum_i \beta_p^i t_p^i \|^2 + \mu \| g_p \odot \beta_p \|^2_2 \right]
\]  

(17)

where \( \beta_p \) contains all the regression coefficients \( \beta_p^i \), the vector \( g_p = [g_p^1, \ldots, g_p^K] \) describes the distance of pixels at position \( p \) with \( g_p^i = |x_p - t_p^i| \), and \( \mu \) is the regularization...
parameter. Apparently, a single pixel can not well describe the local structure. It is essential to further take into account the neighboring pixels in the local neighborhood. We suppose that the details of edges in the target patch can be enhanced by the regression of $K$ selected measured samples, while all the neighboring pixels share the same regression coefficients with the central pixel. The simplest way is to gather the loss functions for all the pixels in the neighborhood. Similar to (17), the objective function can be represented as:

$$
\beta_p = \arg\min_{\beta_p} \sum_{q \in N_p} \left[ x_q - \sum_i \beta_i^p t_q^i \right]^2 + \mu \| g_q \odot \beta_p \|_2^2
$$

where the data term describes the fidelity in the reconstruction and the regularization term encourages the pixels from measured samples with most similar textures to have large regression coefficients. Specifically, the regularization term in (18) can be rewritten as

$$
\sum_{q \in N_p} \| g_q \odot \beta_p \|_2^2 = \| g_N \odot \beta_p \|_2^2
$$

where the element in vector $g_N$ can be represented as $g_{N_p} = \sqrt{\sum_{q \in N_p} (g_q)^2}$. The element $g_{N_p}$ describes the total distance of all the pixels between the target sample and the $i$th measured sample. The pixels from similar measured samples will induce small distances $g_{N_p}$ in the weighting vector $g_N$, which enforces the corresponding coefficients $\beta_i^p$ in vector $\beta_p$.

In (18), the optimization function considers all the pixels in the neighborhood equally. Actually, pixels located near the edge need to be privileged, since they will have much stronger influence for reconstructing the local structure. To solve the above problem, we introduce the steering kernel [44] to adaptively measure the importance of each neighboring pixel. The objective function (18) can be rewritten as a weighted least-square optimization problem:

$$
\beta_p = \arg\min_{\beta_p} \sum_{q \in N_p} \left[ x_q - \sum_i \beta_i^p t_q^i \right]^2 + \mu \| g_q \odot \beta_p \|_2^2
$$

where $w_{pq}^K$ describes the similarity between the pixel $x_p$ and the neighboring pixel $x_q$. It can be represented by:

$$
w_{pq}^K = \frac{\sqrt{\det(C_q)}}{2\pi h^2} \exp \left\{ \frac{-(z_q - z_p)^T C_q (z_q - z_p)}{2h^2} \right\}
$$

where $h$ is the smoothing parameter, the $2 \times 1$ vectors $z_p$ and $z_q$ are the 2-D coordinates of the pixels $x_p$ and $x_q$ respectively. $C_q$ represents the symmetric gradient covariance matrix in the vertical and horizontal directions, which relates to the dominant orientation of gradients in the local neighborhood [44]. To simplify the objective function (19), we can be further represent it as

$$
\beta_p = \arg\min_{\beta_p} \left\{ (x_p - I_p \beta_p)^T W_p^K (x_p - I_p \beta_p) + \mu \| G \beta_p \|_2^2 \right\}
$$

where the vector $x_p = X(N_p)$ represents the concatenation of all the neighboring pixels around position $p$ in the target HR image, $I_p = [I_{H1}(N_p), \ldots, I_{HK}(N_p)]$ represents the $K$ nearest correlative references from the HR training set, the similarity matrix $W_p^K$ and the weighting matrix $G$ are diagonal matrices with $w_{pq}^K(q \in N_p)$ and $g_{i} = \sqrt{\sum_{q \in N_p} w_{pq}^K (x_q - t_q^i)^2}$ on the diagonal respectively.

By differentiating the objective function (21) with respect to $\beta_p$, the solution of the regression coefficient can be derived analytically:

$$
\beta_p = (I_p^T W_p^K I_p + \mu G^T G)^{-1} I_p^T W_p^K P
$$

Once the regression coefficient $\beta_p$ has been estimated, the linear regression (16) can be rewritten as:

$$
x_p = e_p I_p (I_p^T W_p^K I_p + \mu G^T G)^{-1} I_p^T W_p^K P
$$

where $e_p$ is a column vector with the central element (corresponding to position $p$) equal to one and the rest equal to zero. Equation (23) can be simplified as $x_p = a_p^T P$, where the vector $a_p = (e_p I_p (I_p^T W_p^K I_p + \mu G^T G)^{-1} I_p^T W_p^K P)^T$ and each element in $a_p$ represents the corresponding weight of neighboring pixel to estimate the target pixel $x_p$. To regularize the local structure for the whole HR image, we expect to minimize the approximation error for all the pixels, i.e. $\min \sum \| x_p - a_p^T P \|_2^2$. The above minimization term can be described as the following concise form: $\min \| X - AX \|_2^2$, where the matrix $A$ regularizes the relationships of pixels for global optimization. Specifically, $A(p,q)$ represents the contribution of pixel $x_q$ to estimate pixel $x_p$:

$$
A(p,q) = \begin{cases} 
q_i & \text{if } q \in N_p \\
0 & \text{otherwise} 
\end{cases}
$$

Incorporating the above regularization model into (15), the final optimization function can be obtained by:

$$
\omega_p, X = \arg\min_{\omega_p, X} \{ \| Y - DBX \|_2^2 + \eta \| X - AX \|_2^2 \\
+ \alpha \sum_p (\| \phi(R_p X) - \phi(H_p) \omega_p \|_2^2 + \lambda \| d_p \odot \omega_p \|_2^2) \}
$$

s.t. $1^T \omega_p = 1$, $p = 1, 2, \ldots, M$

where the parameter $\eta$ describes the importance of the proposed pixel-based model.

C. Summary of the proposed algorithm

The final output HR face image $X$ can be obtained by solving the above optimization problem (25). However, the objective function is a nonconvex optimization problem. It is difficult to obtain the global optimal solution directly. Fortunately, we can expect to obtain a local optimal solution by an iterative procedure. To deal with the problem in (25), we alternatively optimize over $\omega_p$ and $X$, while keep the other one fixed. If the value of objective $\omega_p$ is fixed, the optimal $X$ can be obtained by minimizing the following function:

$$
f(X) = \| Y - DBX \|_2^2 + \eta \| X - AX \|_2^2 \\
+ \alpha \sum_p (\| \phi(R_p X) - \phi(H_p) \omega_p \|_2^2 + \lambda \| d_p \odot \omega_p \|_2^2)
$$

1057-7149 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
The solution for minimizing (26) can be iteratively addressed by using the gradient descent method. The iterative procedure can be updated by:

\[ X_{n+1} = X_n - \tau \nabla f(X_n) \]  

(27)

where \( \tau \) is the step size of the gradient descent. The gradient of function (26) can be further described as:

\[
\nabla f(X) = 2(DBB)^T (DBX - Y) + 2 \eta (I - A)^T (I - A)X + \alpha \sum_p \nabla J_p(X) \]  

(28)

The kernel function in \( J_p(X) \) projects the image patches from the original HR space to high-dimensional kernel space for better representing the local relationship. In order to simplify (28), we first need to define the concrete expression for kernel function. The Gaussian kernel \( k(x_i, x_j) = \exp(-\sigma ||x_i - x_j||^2_2) \) is utilized in the proposed method, since it can transform the feature to infinite dimension, which helps to seek the nonlinear characteristics. In this case, the gradient function \( \nabla J_p(X) \) can be further described as:

\[
\nabla J_p(X) = \nabla \left\{ \| \phi(R_pX) - \phi(H_p)\omega_p \|^2_2 \right\} 
\]

\[
\quad + \lambda \sum_i (\omega_p^i)^2 \nabla \left\{ \| \phi(R_pX) - \phi(h_p^i) \|^2_2 \right\} 
\]

\[
= -2 \lambda \sum_i (\omega_p^i)^2 \nabla \left\{ \| \phi(R_pX) \|^2_2 \right\} 
\]

\[
\quad - 2 \lambda \sum_i (\omega_p^i)^2 \nabla \left\{ \phi(h_p^i) \right\} 
\]

\[
= -2 \sum_i [\omega_p^i + \lambda (\omega_p^i)^2] \nabla e^{-\sigma \| R_pX - h_p^i \|^2_2} 
\]

\[
= 4 \sigma \sum_i [\omega_p^i + \lambda (\omega_p^i)^2] e^{-\sigma \| R_pX - h_p^i \|^2_2} R_p^T (R_pX - h_p^i) \]  

(29)

When the value of \( X \) is fixed, the optimization of equation (25) reduces to minimize:

\[ J_p(\omega_p) = \{ \| \phi(R_pX) - \phi(H_p)\omega_p \|^2_2 + \lambda \| d_p \odot \omega_p \|^2_2 \} \]

s.t. \( 1^T \omega_p = 1, \quad p = 1, 2, ..., M \)  

(30)

The solution of \( \omega_p \) can be derived from (30) analytically:

\[ \omega_p = (Z^{-1}1)/(1^T Z^{-1}1) \]  

(31)

where the matrix \( Z \) can be obtained by:

\[ Z = (\phi(R_pX)1^T - \phi(H_p)) (\phi(R_pX)1^T - \phi(H_p)) + \lambda (\text{diag}(d_p))^2 \]  

(32)

where \( \text{diag}(\bullet) \) is the diagonalization function for the corresponding vector. Considering the characteristics of Gaussian kernel, each element in the matrix \( Z \) can be concretely represented by:

\[
z_{ij} = \begin{cases} 
1 - e^{-\sigma \| R_pX - h_j \|^2_2} - e^{-\sigma \| R_pX - h_i \|^2_2} & \text{if } i = j \\
+ e^{-\sigma \| h_i - h_j \|^2_2} + \lambda (d_p^i)^2 & \\
1 - e^{-\sigma \| R_pX - h_i \|^2_2} - e^{-\sigma \| R_pX - h_j \|^2_2} & \text{otherwise} 
\end{cases} \]

(33)

The objective function is solved by alternatively optimizing (26) and (30). Thus, it requires an initial estimation for the two objectives in order to begin the iterative process. We first utilize the LR image features to substitute the corresponding features of HR space in (30), which gives a rough estimation for the initial value of \( \omega_p \). Before initialization, the LR images are magnified to the same size of HR ones by bicubic interpolation for obtaining more high-frequency information. We also simply consider the interpolation result of \( Y \) as the initial estimation for the HR face image \( X^{(0)} \).

With the initial value of the two objectives, we can optimize the output HR image \( X \) by minimizing (26). Once this is done, the combination coefficients \( \omega_p \) in each position are obtained by solving (30). We repeat the iterative procedure to update the above objectives until the optimization problem (25) converges to the local minimum. Finally, the optimal solution \( X \) is obtained as the output HR face image. The details of the complete face hallucination procedure are summarized in Algorithm 1.

**Algorithm 1** The proposed face hallucination algorithm.

**Input:** LR testing image \( Y \), LR/HR training images.

1. Initialization: Estimate the initial HR face image \( X^{(0)} \) and combination coefficients \( \{\omega_p^{(0)}\}_{p=1}^M \) on each local position.

2. for \( t=1 \) to \( T \) do

3. \hspace{1em} repeat

4. \hspace{2em} Construct the relationship of local patches in the HR feature space according to (14).

5. \hspace{2em} Learn local edge structure by the weight matrix \( \Lambda^{(t)} \) according to (24).

6. \hspace{2em} Update \( X^{(t)} \) through (26) while fixing the value of \( \{\omega_p^{(t-1)}\}_{p=1}^M \).

7. \hspace{2em} Update \( \{\omega_p^{(t)}\}_{p=1}^M \) through (30) while fixing the value of \( X^{(t)} \).

8. until convergence

9. end for

**Output:** The final output HR face image \( X^* \).

**IV. EXPERIMENTAL RESULTS**

The experiments are conducted on the CAS-PEAL [46] and FERET [47] face databases for demonstrating the performance of the proposed method. For the CAS-PEAL database, we choose a subset that contains 1040 frontal face images with normal expression for experiment. We randomly choose 40 images for testing, while the other 1000 images are utilized for training. For the FERET database, 600 images are selected from the \( ba \), \( bj \) and \( bk \) subsets which correspond to 200 different persons. The \( ba \) subset contains frontal images with regular expression. The \( bj \) subset consists of images with alternative expression, while the \( bk \) subset suffers from illumination variation. Among them, 30 images of 10 individuals are selected for testing, and the other images are served as the training set. All the face images are aligned by the positions of
two eyes and cropped to 128 × 128 pixels. The HR images are blurred using a 7 × 7 Gaussian filter with standard deviation 0.85 and down-sampled to 32 × 32 pixels for producing the LR images. In the experiments, the intensity values of images are converted to the range from 0 (black) to 1 (white) for convenience. We will compare the results of the proposed method with several state-of-art baselines (e.g., LSR [19], WASR [25], LINE [26], SRLSP [28], VDSR [41] and LCGE [37]) in order to validate the effectiveness.

A. Parameter settings

Several parameters are required to be fixed in the proposed method. For the patch-based model, HR patches are set to 12 × 12 in size with four pixels overlapped. The penalty parameter \( \lambda \) in (25) and the parameter \( \sigma \) in Gaussian function is set to 0.001 and 0.01, respectively. In the pixel-based method, the parameter \( \mu \) is fixed to 0.5 and the smoothing parameter \( h \) is selected as 2. We set the regularization parameters \( \eta = 0.05 \) and \( \alpha = 0.01 \) in the optimization function (25). The maximum number of iterations is set to \( T = 10 \).

B. Analysis of Assembling Regularization Models

In the proposed method, two regularization models are assembled together to optimize the target HR image. In this subsection, we aim to discuss the influence of regularization parameters \( \alpha \) and \( \eta \), which are employed to balance the regularization terms in (25). The experiment will also show that the combination of the two regularization models is efficient to enhance the final SR results.

We randomly choose 10 testing images from the CAS-PEAL face database for conducting the experiment. The average PSNR and SSIM [48] values are utilized to evaluate the influence of parameters. We tune parameter \( \eta \) from 0 to 0.1 with step 0.01 and parameter \( \alpha \) from 0 to 0.05 with step 0.005, respectively. Fig. 5 and Fig. 6 show the plots of the average PSNR and SSIM values along with the varying of \( \eta \) and \( \alpha \), respectively. We can observe that the best results are obtained when \( \eta = 0.05 \) and \( \alpha = 0.01 \). When \( \eta = 0 \), it means that the output HR image is merely regularized by the patch-based model. Similarly, the pixel-based model is separately utilized to produce the HR image when \( \alpha = 0 \). Thus, the above experimental results also make a comparison for the effect of the two regularization models. We can observe that the pixel-based model improves the reconstruction quality by 0.42dB in PSNR and 0.0051 in SSIM, while the patch-based model enhances the average quantitative indicators by 0.96dB in PSNR and 0.0095 in SSIM. It illustrates that the combination of two regularization models can produce superior results for face hallucination.

C. Experiments on the CAS-PEAL database

In this subsection, the proposed approach is compared with the state-of-the-art methods to evaluate the effectiveness, which include LSR [19], WASR [25], LINE [26], SRLSP [28], VDSR [41] and LCGE [37]. Some representative hallucinated results are shown in Fig. 7. LSR [19] produces blurring effects on the eyes, nostrils and mouth. The ringing effects also appear around the contour of face image. WASR [25] takes the advantage of sparse representation, which achieves more details than LSR [19] on the face image. However, the staircase noises appear along the margin of face obviously. LINE [26] alleviates the jaggy artifacts near the contour, but the fine details are not well recovered. The textures are also suffered from smoothness around the eyes and mouth. SRLSP [28] enhances the local texture on the face image, but some aliasing artifacts are produced near the areas with high-frequency details. Also, slight blocking artifacts appear on the nose when the image is under high illumination condition. VDSR [41] and LCGE [37] produce face images with more reasonable details by deep learning, whereas the above algorithms sometimes cause noticeable artifacts near the region of eyes. The proposed approach suppresses the noises on the face image and avoids the ringing effects near the contour, while also produces sharper details on the eyeballs and mouth regions. Especially, the proposed method successfully reconstructs the details of two-layer eyelids and pouch in some representative results. Meanwhile, the comparison methods fail to recover the micro high-frequency information on the face image.

In Fig. 8, we further compare the PSNR and SSIM values on the reconstructed HR face images for evaluating the above algorithms. The proposed method achieves the best performance for most testing images. The average PSNR and SSIM values of various methods are shown in Table I. Compared with the second best method [37], the proposed method gains

1. The source codes of LINE [26], SRLSP [28] and LCGE [37] are available on their authors’ webpages.
Fig. 7. SR results on the CAS-PEAL database. (a) Input LR image. (b) LSR [19]. (c) WASR [25]. (d) LINE [26]. (e) SRLSP [28]. (f) VDSR [41]. (g) LCGE [37]. (h) The proposed method. (i) Original HR image. (The apparent distinctions between the compared algorithms and the proposed one are marked in the red rectangle. Please refer to the electronic version and zoom in for better comparison.)

Fig. 8. PSNR and SSIM values on the CAS-PEAL database. (a) PSNR. (b) SSIM.

The average PSNR and SSIM values for CAS-PEAL database.

<table>
<thead>
<tr>
<th></th>
<th>LSR</th>
<th>WASR</th>
<th>LINE</th>
<th>SRLSP</th>
<th>VDSR</th>
<th>LCGE</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>29.67</td>
<td>30.69</td>
<td>30.95</td>
<td>31.06</td>
<td>31.12</td>
<td>31.13</td>
<td>31.73</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9177</td>
<td>0.9290</td>
<td>0.9317</td>
<td>0.9333</td>
<td>0.9345</td>
<td>0.9350</td>
<td>0.9404</td>
</tr>
</tbody>
</table>

Table I

The average PSNR and SSIM values for FERET database.

<table>
<thead>
<tr>
<th></th>
<th>LSR</th>
<th>WASR</th>
<th>LINE</th>
<th>SRLSP</th>
<th>VDSR</th>
<th>LCGE</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>31.74</td>
<td>32.19</td>
<td>32.19</td>
<td>32.34</td>
<td>32.45</td>
<td>32.59</td>
<td>32.83</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.8941</td>
<td>0.9005</td>
<td>0.8995</td>
<td>0.9002</td>
<td>0.9088</td>
<td>0.9057</td>
<td>0.9088</td>
</tr>
</tbody>
</table>

Table II

the improvement of 0.60dB in PSNR and 0.0054 in SSIM. The experimental results further illustrate the advantage of the proposed method in quantitative comparison.

D. Experiments on the FERET database

We also conduct the experiments on the FERET database in order to show the robustness of the proposed method. The testing face images are acquired under significantly different illuminations and face expressions. It is utilized to evaluate the proposed algorithm under various conditions. Some comparison results are presented in Fig. 9. The proposed approach produces best visual quality, both in terms of flat textures and details on high-frequency areas (e.g. eyes, teeth and mouth), while other approaches produce some noticeable artifacts on the face image. Fig. 10 presents the corresponding PSNR and SSIM values for the comparison of different methods. The average quantitative comparisons are also shown in Table II. Compared with LCGE [37], our method achieves an increase of 0.24dB in PSNR and 0.0031 in SSIM. The experimental results demonstrate that our method has the ability to deal with the variations of illumination and facial expression.
Fig. 9. SR results on the FERET database. (a) Input LR image. (b) LSR [19]. (c) WASR [25]. (d) LINE [26]. (e) SRLSP [28]. (f) VDSR [41]. (g) LCGE [37]. (h) The proposed method. (i) Original HR image. (The apparent distinctions between the compared algorithms and the proposed one are marked in the red rectangle. Please refer to the electronic version and zoom in for better comparison.)

Fig. 12. SR results on the real-world image. (a) Input LR image. (b) LSR [19]. (c) WASR [25]. (d) LINE [26]. (e) SRLSP [28]. (f) VDSR [41]. (g) LCGE [37]. (h) The proposed method. (The apparent distinctions between the compared algorithms and the proposed one are marked in the red rectangle. Please refer to the electronic version and zoom in for better comparison.)

Fig. 11. The real-world image from Internet.

E. Experiments on the real-world images

In this subsection, we conduct the experiments on the real-world images for comparison. Fig. 11 shows the real-world image which is collected from the Internet. The marked human faces are hallucinated in the experiment. The LR faces are manually aligned through the positions of two eyes and standardized to the size of $32 \times 32$. Since individuals in the image are all from America, we utilize the FERET face database as the training set for reconstructing the HR face image. Fig. 12 presents the hallucinated results for various methods. From the experimental results, we can see that the methods of LSR [19], WASR [25] and LINE [26] are very sensitive to noises. Lots of aliasing artifacts appear on the...
reconstructed face image, especially on the regions containing complex details. Furthermore, it seems that the textures of eyes are mixed with the edge of eyelid in some SR results. The methods of [28] [41] and [37] have the ability to improve the noisy effects on the SR results. However, the reconstructed results are somewhat suffered from artifacts. It is apparent that the details around the area of eyes are not very clear. The proposed method is able to produce very reasonable results. It suppresses the noises on the face image and recovers fine details for the SR results, which achieves better performance than other methods.

In order to further validate the effectiveness of various algorithms in real-world environment, we conduct an extensive experiment on the CelebA database [49]. The images of the CelebA dataset are all obtained from the Internet, which includes different lighting conditions and facial expressions from various environments. In the experiment, we select 1000 face images with near-frontal pose as training samples, while another 50 near-frontal face images are served as the testing set. The HR face regions are cropped and resized to 128×128 pixels. Then the 32×32 LR face images are obtained by blurring and down-sampling procedures. Fig. 13 presents some representative SR results for comparison. The proposed method produces fine details especially on the regions of eyes and mouth. The PSNR and SSIM values for all the testing images are shown in Fig. 14, while the average PSNR and SSIM values are also shown in Table III. Compared with other algorithms, the proposed method achieves the best performance in both visual quality and quantitative indicator. The experimental results illustrate that the proposed method has the capability to deal with face image in real-world environment.

### TABLE III

<table>
<thead>
<tr>
<th>Method</th>
<th>LSR</th>
<th>WASR</th>
<th>LINE</th>
<th>SRLSP</th>
<th>VDSR</th>
<th>LCGE</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR(dB)</td>
<td>33.23</td>
<td>33.55</td>
<td>33.57</td>
<td>33.69</td>
<td>33.57</td>
<td>33.51</td>
<td><strong>34.25</strong></td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9167</td>
<td>0.9207</td>
<td>0.9208</td>
<td>0.9227</td>
<td>0.9227</td>
<td>0.9220</td>
<td><strong>0.9297</strong></td>
</tr>
</tbody>
</table>

In the former experiment, we select 40 images from the database for testing, while all the remaining 1000 images are served as the training set. To study the relationship between training set size and reconstruction quality, we decreases the size of training set from 1000 to 200 with step 200. Meanwhile, the 40 testing images are always fixed in the experiment for further comparison. We then perform the face hallucination experiments and compare the reconstruction results on training data with various amounts of samples. Some representative hallucinated images are shown in Fig. 15. The proposed method preserves sharp details and produces reasonable HR images with all the training sets. The change of visual quality is very slight along with the variation of training set size.

### TABLE IV

<table>
<thead>
<tr>
<th>Size</th>
<th>200</th>
<th>400</th>
<th>600</th>
<th>800</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR(dB)</td>
<td>30.82</td>
<td>31.37</td>
<td>31.54</td>
<td>31.66</td>
<td>31.73</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9315</td>
<td>0.9370</td>
<td>0.9391</td>
<td>0.9399</td>
<td>0.9404</td>
</tr>
</tbody>
</table>

Fig. 16 presents the PSNR and SSIM values for the 40 testing images with various training set. The quantitative indicators gradually increase together with the amount of training samples. Especially, the proposed method achieves fine results even when the training set only contains 200 image samples. The experimental results show that the proposed algorithm is insensitive to the size of training set.

The average PSNR and SSIM values are also presented in Table IV. The advantage of the proposed method is remarkable. Compared with the state-of-the-art methods shown in Table I (with 1000 training samples), it achieves superior quantitative indicators with only 400 samples. The experimental results further demonstrate that the proposed algorithm has the capability of producing satisfactory results with a small number of training samples.

### G. Comparison on Resource-limited Training Set

Recently, plenty of algorithms in computer vision area aim to train complex models from a mass amount of data. Though the increase of training data really induces the improvement of performance, we have to deal with the lack of data in the real-world application when the resource of training set is limited.

---

Fig. 14. PSNR and SSIM values on the CelebA database. (a) PSNR. (b) SSIM.

Fig. 16. PSNR and SSIM values with various amounts of training samples. (a) PSNR. (b) SSIM.
Fig. 13. SR results on the CelebA database. (a) Input LR image. (b) LSR [19]. (c) WASR [25]. (d) LINE [26]. (e) SRLSP [28]. (f) VDSR [41]. (g) LCGE [37]. (h) The proposed method. (i) Original HR image. (The apparent distinctions between the compared algorithms and the proposed one are marked in the red rectangle. Please refer to the electronic version and zoom in for better comparison.)

Fig. 15. SR results on the CAS-PEAL database with various amounts of training samples. (a) Input LR image. (b) 200 training images. (c) 400 training images. (d) 600 training images. (e) 800 training images. (f) 1000 training images. (g) Original HR image. (Please refer to the electronic version and zoom in for better comparison.)

Fig. 17. SR results on the CAS-PEAL database with only 100 training samples. (a) Input LR image. (b) LSR [19]. (c) WASR [25]. (d) LINE [26]. (e) SRLSP [28]. (f) VDSR [41]. (g) LCGE [37]. (h) The proposed method. (i) Original HR image. (The apparent distinctions between the compared algorithms and the proposed one are marked in the red rectangle. Please refer to the electronic version and zoom in for better comparison.)
Thus, we devise an experiment on the CAS-PEAL database for studying the robustness of various algorithms. We fix the LR testing images as previously mentioned, while only 100 images are randomly selected from the whole data set for training. It means that the amount of training images decreases from 1000 to 100. Fig. 17 shows some representative SR results for various algorithms. With resource-limited training data, the visual quality of hallucinated image is apparently degraded for all the methods. It can be seen that serious artifacts have been produced on the output HR images. Though some noisy effects appear on the results of the proposed method, it consistently preserves the ability to generate plausible high-frequency details, especially on the region of eyes and mouth. Meanwhile, the algorithms [19], [25], [26], [28], [41] tend to induce ringing effects and unexpected visual artifacts on the face image. The noisy and ringing effects can be suppressed by [37], but the results of [37] seem to suffer from smoothness and lose the local details. We further compare the PSNR and SSIM values for all the testing images in Fig. 18. The average PSNR and SSIM values are also shown in Table V. The methods [41] [37] and the proposed method achieve higher quantitative indicators when compared with other approaches. The proposed method also produces more vivid details than [41] [37], which is important to the discriminability of human visual system. The experimental results further illustrate that the proposed algorithm has the advantage of producing superior results with resource-limited training data.

![Fig. 18. PSNR and SSIM values on the resource-limited training set. (a) PSNR. (b) SSIM.](image)

### Table V

<table>
<thead>
<tr>
<th>LSR</th>
<th>WASR</th>
<th>LINE</th>
<th>SRLSP</th>
<th>VDSR</th>
<th>LCGE</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>29.01</td>
<td>29.39</td>
<td>29.39</td>
<td>29.35</td>
<td>30.08</td>
<td>30.15</td>
<td>30.19</td>
</tr>
<tr>
<td>0.8997</td>
<td>0.9072</td>
<td>0.9065</td>
<td>0.9058</td>
<td>0.9250</td>
<td>0.9230</td>
<td>0.9211</td>
</tr>
</tbody>
</table>

We also compare the computational time of the proposed algorithm with other methods. LSR [19], WASR [25], LINE [26], SRLSP [28], VDSR [41] and the proposed algorithm are implemented on a computer of 3.2 GHz CPU and 8G memory using MATLAB. VDSR [41] and LCGE [37] are conducted on a single Tesla K80 GPU using Caffe, while the enhancement part in [37] is performed using C++ on CPU. The deep learning based methods (VDSR [41] and LCGE [37]) respectively require about eight hours and two days to train the deep network on GPU, while other methods do not need the pre-training procedure. The average running time of LSR [19], WASR [25], LINE [26], SRLSP [28], VDSR [41], LCGE [37] and the proposed method are 10.6 s, 24.5s, 17.1s, 9.3s, 2.1s, 72.8s and 35.5s, respectively. Though the proposed method does not present the superiority in computational time, it has the ability to produce superior results with reasonable time cost. The proposed method also has the advantage to produce HR results without the time-consuming training process. Furthermore, it is feasible to improve the efficiency of the proposed method by reducing the number of iterations. According to Fig. 19, the proposed method achieves 31.47dB in PSNR and 0.9387 in SSIM after the second iteration. In such case, the computational cost will reduce to about one-fifth of the current time cost, which can then beat most of the existing methods.

### I. Effectiveness of utilizing RKHS

In the following, we will design an experiment to show the effectiveness of using RKHS in the reconstruction. In this experiment, we abandon to project the patches into RKHS and compare the experimental results with the proposed method. Specifically, the reconstruction model without using RKHS can be written by setting $\phi(x) = x$ in (25). The details of deduction are shown in the Supplemental Material. The experiments are conducted on the CAS-PEAL database, while the 40 testing images are fixed in the experiments for comparing the visual quality and quantitative indicators. Some representative hallucinated results are presented in Fig. 20. The proposed algorithm obtains vivid results for the details see that the proposed algorithm achieves superior results with the increase of iteration number, while the PSNR and SSIM measures become stable after 10 iterations. Due to the balance of reconstruction quality and efficiency, the maximum number of iterations is set to 10 in the proposed method.
of two-layer eyelids. The PSNR and SSIM values for all the testing images are shown in Fig. 21. It is apparent that the reconstruction in the RKHS produces better results than the algorithm in the original HR space. The average PSNR and SSIM values of all the testing images are 31.47dB and 0.9385 for the reconstruction algorithm without using RKHS. Compared with the results in Table I, the proposed method achieves the improvement of 0.26dB in PSNR and 0.0019 in SSIM, respectively. The experimental results demonstrate that the reconstruction in the RKHS is effective to obtain superior HR images for face hallucination.

![Fig. 20. Comparison for the effectiveness of utilizing RKHS. (a) Without RKHS. (b) The proposed method. (Please refer to the electronic version and zoom in for better comparison.)](image)

![Fig. 21. PSNR and SSIM values for the experimental results. (a) PSNR. (b) SSIM.](image)

**J. Effectiveness of the patch-based and pixel-based models**

In this subsection, we respectively give the reconstruction results for the patch-based and pixel-based models to show the effectiveness of each regularization term. The experiments are conducted on the CAS-PEAL database, while the results of utilizing various regularization models are obtained by alternately adjusting the parameter $\eta$ and $\alpha$ in (25). When we set $\eta = 0$, it means that the output HR image is produced by the patch-based regularization model. When the parameter $\alpha$ is set to 0, the pixel-based model is utilized to obtain the output HR image. Some representative SR results are shown in Fig. 22. The PSNR and SSIM values of all the testing images are presented in Fig. 23. The algorithm with separated patch-based regularization term achieves the average quantitative indicators as 31.36dB in PSNR and 0.9355 in SSIM, while the average PSNR and SSIM values are 30.81dB and 0.9314 for the algorithm with separated pixel-based regularization term. According to the experimental results, the patch-based model produces superior results than the pixel-based one when they are utilized separately. However, the performance of the combined regularization terms constantly outperforms the results of two separated regularization terms as shown in Fig. 23. Thus, both of the two regularization models play an important role in the proposed method.

![Fig. 22. Face hallucination results for utilizing various regularization models. (a) Pixel-based regularization model. (b) Patch-based regularization model. (c) The proposed method. (Please refer to the electronic version and zoom in for better comparison.)](image)

![Fig. 23. Comparison for the patch-based and pixel-based models. (a) PSNR. (b) SSIM.](image)

**K. The performance of face recognition**

In this subsection, we evaluate the performance of the above face hallucination algorithms in face recognition application. The experiments are conducted on the Multi-PIE database [50]. We select a subset of session 04 which contains frontal face images with neutral expression under 20 different illumination conditions (Camera: 05_1, Recording Number: 01). For each individual, five face images are randomly selected as the testing set, while the rest of images are considered as the training set for the experiments of recognition. The HR face images are cropped and normalized to $64 \times 64$ pixels, while the corresponding LR images are down-sampled to the size of $16 \times 16$ pixels for performing the experiments in a low resolution. To evaluate the effect of face hallucination, the LR testing images are hallucinated to the resolution of $64 \times 64$ pixels using the training set. The face recognition experiments are then performed on the hallucinated HR face images. Since the experiments focus on the comparison of recognition ability for various face hallucination methods, we employ PCA to obtain the eigenfaces and utilize the 1NN classifier for recognition. The simplest recognition algorithm is helpful to demonstrate that the improvement of recognition rate owes to the face hallucination methods rather than other
TABLE VI
THE FIRST-RANK RECOGNITION RESULTS FOR VARIOUS FACE HALLUCINATION ALGORITHMS

<table>
<thead>
<tr>
<th></th>
<th>Bicubic</th>
<th>LSR</th>
<th>WASR</th>
<th>LINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>85.57%</td>
<td>68.14%</td>
<td>85.96%</td>
<td>80.09%</td>
</tr>
<tr>
<td>LSLSP</td>
<td>81.56%</td>
<td>81.56%</td>
<td>79.13%</td>
<td>81.90%</td>
</tr>
</tbody>
</table>

sophisticated classification algorithms. To better compare the experimental performance, we also provide the recognition rates on the degraded LR images and original HR images as the baselines. The recognition results of various algorithms are shown in Table VI. The recognition rate of the proposed method is 81.90%. It achieves the highest recognition rate when compared with other face hallucination approaches. It is about 13% higher than the result on the degraded LR images without face hallucination. We also notice that the bicubic interpolation slightly decreases the final recognition rate because it produces artificial effects on the reconstructed HR images. The experiments demonstrate that the proposed algorithm has the ability to reconstruct faithful details and improve the recognition performance.

V. CONCLUSION

In this paper, we propose two novel regularization models to deal with LR face hallucination problem. The patch-based model regularizes the relationship between target patch and training patches in the HR space, which is effective to recover local texture. Different from the previous methods which build the regularization in the LR space, it avoids the difficulty of preserving local geometry consistency. Furthermore, the HR patches are projected into the RKHS, which helps to seek the nonlinear relationship in reconstruction. The pixel-based model is devised to compensate local details, especially on the regions of eyes, nostrils and mouth. It privileges the pixels along the dominant orientation of structure, which is valid to reconstruct the local edge structure. The whole HR image can be finally optimized by combining the above two regularization models. Experimental results illustrate that the proposed method produces superior results in comparison with state-of-the-art baselines under various conditions. Future work will focus on the improvement of the proposed method for solving multiview face hallucination problem with large pose variation.

REFERENCES


