Face image super-resolution through locality-induced support regression

Junjun Jiang, Ruimin Hu, Chao Liang, Zhen Han, Chunjie Zhang

A R T I C L E   I N F O

Article history:
Received 7 June 2013
Received in revised form
17 December 2013
Accepted 17 February 2014
Available online 24 February 2014

Keywords:
Super-resolution
Face image
Support regression
Manifold learning

A B S T R A C T

In this paper we propose a novel face image super-resolution (SR) method named Locality-induced Support Regression (LiSR). Given a low-resolution (LR) input patch, we learn a mapping function between the local support LR and high-resolution (HR) patch pairs to predict its HR version. The support can be obtained from the LR or HR patch manifolds, which leads to two varieties of LiSR, namely LR patch guided LiSR (LR-LiSR) and HR patch guided LiSR (HR-LiSR). LR-LiSR directly learns the mapping function between local support LR/HR patch pairs given an input LR patch. As for HR-LiSR, the support and a mapping function will be iteratively learned to update the target HR patch. The key advantages of our proposed framework are two-fold: (1) the strong regularization of “same representation” of prior work [1,2] is relaxed to the same support, and hence much flexibility can be given to the learned mapping function; (2) we define the support in the LR or HR patch manifold space by incorporating the locality constraint, which can well preserve the manifold structure of the training set. Experimental results reported on both simulated LR face images and real-world datasets demonstrate the effectiveness of the proposed method.

1. Introduction

In most digital imaging applications, such as medical diagnosis, remote sensing, and intelligence surveillance, images of high signal-noise-ratio (SNR) are ideal data source for high-level visual tasks such as object recognition, digital entertainment, and remote sensing imaging. To increase SNR, some researchers develop various denoising methods to erase additional noisy signal attached to the original image [3,4], while others aim to enhance the image quality through the SR technique, which is also the focus of this paper. In order to increase the resolution of an electronic imaging system, one of the most straightforward ways is to increase the sensor density by reducing the sensor size. However, with the increase of sensor density, both hardware cost and shot noise will inevitably increase [4]. Therefore, the hardware limitation on the size of the imaging sensors restricts the resolution of captured images [5].

Another promising way to address this problem is to use signal processing techniques to post-process captured images, aiming to break the hardware limitations of the electronic imaging system. These techniques are specifically referred as image SR [6,7].

SR, a.k.a. image hallucination, is the technique that constructs a HR image from a LR image sequence or a single LR image, thereby increasing the high frequency components and removing the degradations caused by the LR imaging device. In the past decade, this SR technique has received wide research interests, and a number of approaches have been proposed. These methods can be roughly categorized into three categories, i.e., interpolation-based methods [8–10], reconstruction-based methods [11–14], and learning-based methods [15–18].
The interpolation-based methods are intuitive and highly efficient, such as Bilinear, Bicubic, and other resampling methods [8–10]. However, they usually end up with unsatisfactory performance in the case of larger magnification. The reconstruction-based methods are established on the imaging degradation model, and their basic idea is to combine the non-redundant information contained in multiple LR images to generate a HR one. However, as the magnification factor increases, the reconstruction constraints, which are normally combined with some form of smoothness prior to regularize the solution, provide less and less useful information [16]. As reported in Lin’s work [19], under practical conditions, the limit magnification factor of reconstruction-based methods is found to be 1.6 if the denoising and registration are not good enough. In addition to reconstruction constraint, the learning-based methods, also referred as example-based methods, infer the high-frequency details lost in the input LR image from a training set of LR and HR image pairs, i.e., the relationship between LR images and the corresponding HR ones can be used to estimate the missing HR frequency details in the input LR image. In this paper, we focus on the learning-based method for its stronger SR capability, especially when the magnification is large, e.g., a factor of more than two times [16].

Besides the classification mentioned above, SR can also be classified into general SR and domain-specific SR, according to the type of applied LR images. Different from the general SR methods where the prior is learned from general images, the domain-specific SR methods learn the domain-specific degradation models from some special objects. Obviously, for the domain-specific SR problem, such as human face image SR, the SR performance is better than that of general SR methods, which do not consider the specific properties of human faces, and this point will be verified in the experiments. An comprehensive overview of current advances in face image SR is given in [20].

In this paper, we focus on the learning-based face image SR. Traditional methods assume that LR and HR image spaces share the same local geometry structure, and use the reconstruction relationship obtained from LR space to recover the HR image (patches). However, due to the “one-to-many” mapping between LR and HR images [21] (for natural images different HR images can result in the same LR image when blurred and down-sampled), the neighborhood relationship in the LR space cannot be directly conveyed to the HR space. In this paper, we propose to learn the regression function, instead of neighborhood relationship, between the support LR and HR patch pairs, and the support is defined in the HR space, instead of in the LR space, to avoid the “one-to-many” mapping problem in the degradation process. In particular, given a LR input face image and LR/HR training pairs, we first divide all the images into patches according to the position using the same dividing strategy. For each patch in the input LR face, we try to learn a mapping function between the support LR/HR patch pairs of the same position to estimate the HR patch. We define the support by incorporating the locality constraint of manifold structure. Therefore, the learned regression model cannot only ensure the data fidelity of LR and HR mapping, but also explore and preserve the geometrical structure of the LR/HR training sets. Acquiring all the super-resolved HR patches, we integrate them according to the position to generate the target HR face image.

1.1. Related work

Baker and Kanade [15] developed a learning-based face image SR method named “face hallucination” to infer the HR face image from an input LR one based on face priors, and this is the first SR method targeted at face images specially. Liu et al. [17] proposed a two-step statistical modeling approach to integrate a global parametric model and a local nonparametric model for face image SR. Following the pioneering work of [15] and [17], learning-based face image SR methods draw enormous attention recently. Roughly speaking, given a LR observation and the LR/HR training pairs, there are two classes of techniques [21], one being coding-based methods [1,2,22–25] and the other regression-based methods [18,26–28], for super-resolving LR images.

**Coding-based methods:** Based on the assumption that LR and HR patches form manifolds with local similar geometry in two distinct spaces, the coding-based methods firstly explore the relation between the LR training samples, and then directly preserve this explored relation, e.g., the same coding coefficients (or called same representation), for the HR training samples, thus inferring the target HR image. As shown in Fig. 1, the input LR patch is first encoded on the LR training samples, and then the corresponding HR training samples and the same coding coefficients are used to construct the target HR patch.

![Block diagram of coding-based face image SR method.](image)

**Fig. 1.** Block diagram of coding-based face image SR method.
The most representative coding-based method is the Neighbor Embedding (NE) SR method proposed by Chang et al. [1]. The algorithm selected K best candidates from the LR training set, and used the corresponding HR training patches and the same representation obtained in the LR training set to linearly reconstruct the target HR patch. Nevertheless, using a fixed K number of neighbors for reconstruction may lead to blurred edges, due to over- or under-fitting. To alleviate this problem, Yang et al. [2] employed a Sparse Coding (SC) method to adaptively choose the most representative samples for image SR. Similarly, they first obtained the coefficients (i.e., the outcome of the sparse coding process) in the LR dictionary given an input LR patch, and then the super-resolved HR patch is reconstructed by replacing the LR dictionary with its HR counterpart. Inspired by the local coordinate coding [29,30], in our previous work [25], we further improved the face image SR methods by giving different freedom for LR training patches proportional to its similarity to the input LR patch. As far as we know, this Locality constraint Representation (LcR) based face image SR method obtains the best performance reported in the literatures.

However, due to “one-to-many” mapping between LR and HR images [31], the above assumption (LR and HR patches share the same neighboring relation) is sometimes too idealistic to meet in the practical applications. Although many approaches have been proposed to mitigate the negative impact of this problem [32,33], their essential is still trying to preserve the relation of the LR training samples for HR ones or project LR and HR images to a common space and then perform SR [1].

Regression-based methods: Another class of learning based SR methods are regression-based, which pose the problem of SR as a regression problem and directly predict the HR output via a set of learned mapping functions. Without any additional assumption, the regression-based method directly learns the mapping function between the LR and HR training samples, and then the learned mapping function is applied to estimate the HR image given a LR observation as shown in Fig. 2.

For example, Tappen et al. [18] posed the problem of estimating the high-frequency details as a multiple linear regression problem on a clustered example database and resolved the resulting multiple candidate HR outputs by imposing a prior on natural images. Ni and Nguyen [26] exploited support vector regression to solve the regression problem of the high frequency details in the frequency domain. In [27], Kim and Kwon proposed a Kernel Ridge Regression (KRR) SR method. To remove the blurring and ringing effects around strong edges because of the regression, a prior image model which takes into account the discontinuity property of images was used for post-processing. Most recently, Tang et al. [34] learned the linear regression relationship in a special feature space, i.e., sparse coding space, and proposed a Greedy Sparse Coding Regression (GSCR) for general image. Taking the position prior of human face, Huang and Wu proposed a frontal facial image SR approach by using multiple Local Linear Transformations (LLT) [28]. It learns the linear transformations between LR and HR training patches of the same position and the SR processing only requires simple matrix multiplications, thus is very efficient and applicable to resource-limited systems. LLT is essentially a global regression model, which reduces the SR to a projection of each input LR patch into the HR space by multiplication with a pre-computed matrix. It is however a global solution and cannot be adaptively adjusted according to the LR observation, thus the generalization ability is limited.

Instead of considering the whole training set, these kernel regression models [26,27] consider to learn the mapping function on the support according to the observed LR patch, thus much more flexible. However, the support is defined in the LR space, which cannot reflect the real geometry of samples due to the inconsistency between the LR and HR space.

1.2. Contribution

From the discussion presented above, we can see that the inconsistency between the LR and HR space is the biggest obstacle for these coding-based and regression based approaches. To this end, we directly learn the mapping function to recover the target HR patch from the given LR patch, without the “same representation” assumption widely as in coding-based methods. In order to exploit the geometry of the HR space, we further propose an iteration framework. In particular, through incorporating the locality constraint of manifold structure, the support is defined by the indices of K-nearest neighbors between the input LR patch and the LR (HR) training patches, which lead to the LR (HR) patch guided support generation strategy. For the HR patch guided support generation strategy, as we could not obtain the HR version of one LR input patch beforehand, we design an iterative
schema to update the support and learn a new mapping function from previous support LR/HR patch pairs. Compared with previous face image SR work, our contribution can be summarized as follows:

- Compared with those coding-based methods [1,2,22–25], which use the strong regularization of “same representation” for learning, we relax the “same representation” to “same support”, giving more flexibility to the learned mapping function.
- Instead of learning a global mapping function from the entire training samples as in regression-based methods [18,26–28], we design to learn the specific mapping function for each observation (one input LR patch) from its support LR/HR patch pairs, and thus the learned mapping function can be tuned towards a specific input LR patch.
- Compared with those regression-based methods [18,26–28], which ignore the geometry of the HR patch space, we define the support set by the geometry of the HR patch space and use the geometry to regularize the mapping function. With an iterative optimization technology, the proposed method can produce more detailed face features step by step.

Note that we previously proposed a regression-based method, namely Manifold regularized Sparse Support Regression (MSSR), for general image SR [35]. Although MSSR and the proposed method all try to learn the mapping relationship between the LR patches and HR ones on the support, they have some essential differences. In particular, MSSR defined the support set of the input LR patch with these LR training patches with non-zero sparse coding coefficients. However, due to the fact that many HR images may correspond to one LR image, the neighborhood relationship of the LR space cannot reflect the truth. To this end, instead of defining the support set in the LR image patch space as in MSSR, we obtained the support set in the HR image patch space (using the estimated HR patch and leading to HR-LiSR), whose geometry is much more credible and discriminant than that of the LR image patch space [31]. Since the target HR patch is unknown in advance, we formulate the target HR patch SR as an iterative optimization problem (while the support set and the mapping function are learned in one time in the MSSR method). Therefore, the super-resolved results can be refined step by step. In addition, MSSR aims at super-resolving the general scene and does not consider the prior of face image, while LiSR is specially designed for facial image. Through incorporating the face position prior (all face images have similar structures and the patches at the same site are highly related once we align the face images according to the positions of two eyes), LiSR establishes model for each position patch but for the entail face image, thus leading to more flexible SR framework.

1.3. Organization of this paper

The rest of the paper is organized as follows. The details of the proposed LiSR approach are presented in Section 2. Comparative results are reported in Section 3 and a brief discussion is given in Section 4. Finally, we give concluding remarks and future prospects in Section 5.

2. The proposed algorithm

In this section, we present the detailed procedure of the proposed approach. We begin with the terms and notations. As stated, the problem of face image SR is formulated as the inference of the HR face image \( y_t \) from an input LR face image \( x_t \), given the training sets of HR and LR face images, \( Y = [y_m]_{m=1}^{M} \) and \( X = [x_m]_{m=1}^{M} \), where \( M \) denotes the size of the training sets. As in many face image SR approaches [23–25], we represent each face by image patches. Therefore, each face image mentioned above is divided into \( N \) small overlapping patch sets \( \{y_{m(i,j)}\} \leq i \leq U, 1 \leq j \leq V \}_{m=1}^{M} \) and \( \{x_{m(i,j)}\} \leq i \leq U, 1 \leq j \leq V \}_{m=1}^{M} \), the patch number of each face image is calculated by \( N = UV \), \( U \) denotes the patch number in every column, \( V \) denotes the patch number in every row, and the term \( (ij) \) indicates the coordinate in the patch coordinate system \( \text{o}_{uv} \), as illustrated in Fig. 3.

For one input LR face image denoted in patches as \( \{x_t(i,j)\} \leq i \leq U, 1 \leq j \leq V \), the face image SR approaches super-resolve each input LR patch \( x_t(i,j) \) to obtain its HR version \( y_t(i,j) \). Concatenating and integrating all the super-resolved HR patches \( \{y_t(i,j)\} \leq i \leq U, 1 \leq j \leq V \) according to their corresponding positions, we can obtain a face image, which is the target HR face image \( y_t \).

Specially, the coding-based approaches encode the input LR patch on the LR training patches of the same position by a linear combination of neighbors, thus obtaining the coding coefficients:

\[
\hat{\theta} = \arg\min \{\|x_t(i,j) - X(i,j)\theta\|^2 + \lambda E(\theta)\},
\]

where \( X(i,j) \) is a matrix with its columns being training patches, \( X(i,j) = [x_t(i,1), x_t(i,2), \ldots, x_t(i,j)] \), \( E(\theta) \) is a prior of the coding coefficients, which enforces the special

![Fig. 3. Dividing a face into \( N=UV \) patches. The term \((ij)\) indicates the coordinate of one patch in the patch coordinate system \( o_{uv} \). patch_size and overlap denote the side pixels of one square patch and the overlap pixels between patches respectively.](image)
properties of the resulting image, such as smooth prior, sparsity prior, or locality prior, and $\lambda$ is positive scalar that controls the balance between the reconstruction error and the prior information. Acquiring the coding coefficients $\hat{\vartheta}$, these coding-based approaches directly transform $\vartheta$ to the HR training patches, $Y(i,j) = [Y_1(i,j), Y_2(i,j), \ldots, Y_M(i,j)]$, of the same position to obtain the target HR patch:

$$y_t(i,j) = Y(i,j)\hat{\vartheta}.$$  

(2)

The coding-based approaches all assume that each pair of HR and LR patches has the “same representation”. However, due to the projection from the LR image to the HR image is “one-to-many” mapping, this strong regularization of same representation is not always the truth in practical applications.

In the regression-based approaches, the assumption is relaxed for LR/HR image mapping, it directly models the regression relationship between the LR training patches and the HR training patches, $f : x \rightarrow y$, and then utilizes the learned regression model $f$ to predict the target HR patch $y_t(i,j)$ given an input LR patch $x_t(i,j)$,

$$y_t(i,j) = f(x_t(i,j)).$$  

(3)

Without strong prior on the relationship between LR and HR training images, the regression-based approaches are much more flexible than that of these coding-based approaches. Note that image patches have regular structures where accurate estimation of pixel values via regression is possible.

Inspired by these ideas, in this paper, we proposed a novel local linear regression model (LiSR) to model the relationship between LR and HR patches. Specially, the support information can be obtained by two strategies, the LR patch guided support generation strategy and the HR patch guided support generation strategy, and these will lead to two varieties of LiSR, LR patch guided LiSR (LR-LiSR) and HR patch guided LiSR (HR-LiSR). In the following section, we will give the details of these two models.

2.1. LR-LiSR

In contrast to those coding-based methods [1,2,22–25], the “same representation” of a pair of LR and HR patch is relaxed for having the same support in LiSR. As a result, LiSR directly models the regression relationship between LR and HR patches on the support. It is flexible in using the information of local training samples rather than the entire training samples as in [28]. In the following, the LR patch guided LiSR model will be given.

Suppose that we are given a training set with LR and HR patch pairs of the position $(i,j)$ as in the following:

$$\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}.$$  

(4)

Here, we drop the position index $(i,j)$ for convenience from now on. For an unseen LR patch $x_t$, we try to learn a mapping function, $f(x, P) = Px$, from LR patches to HR patches on the support to minimize the following regularized cost function:

$$O_{LR-LiSR} = \sum_{m \in S_t} (Px_m - y_m)^2 + \alpha\|P\|^2_F,$$  

(5)

where $P$ is the mapping matrix to be learned, $\|P\|^2_F$ is a smooth term, $\alpha$ is a small positive parameter (such as $\alpha = 10^{-5}$), and $S_t$ is the support of the input LR patch on the LR training set. Now the question becomes: how to estimate the support $S_t$ for the input LR patch.

Recently, manifold learning theory states that although data are physically represented in a high-dimensional space, they often lie on a manifold which has a much smaller intrinsic dimensionality. The core idea in manifold learning is the local geometrical estimation and preservation. And this data locality has been widely utilized in many computer vision applications such as clustering, dimension reduction, and image classification. In this paper, we propose to take advantage of the local manifold geometric structure to learn the regression function between LR and HR training patches. The main idea is to locally map the LR patch on the LR patch manifold into HR patch manifold. In particular, the neighbor information is applied to guide the estimation of support. Inspired by this, we define the support as follows:

$$S_t = \text{supp}(\text{dist}_{1:k}(\cdot)), $$  

(6)

where $\text{dist}_{1:k}$ refers to the smallest $K$ entries of the $\text{dist}_{1}$, thus $\text{supp}(\text{dist}_{1:k})$ is a set composing the indexes of these smallest $K$ entries in $\text{dist}_{1}$, and $\text{dist}_{1} \in \mathbb{R}^M$ is the measurement of squared distances between $x_t$ and the LR training patches $X$. Specifically,

$$\text{dist}_{1:m} = \|x_t - x_m\|^2, \quad 1 \leq m \leq M,$$  

(7)

where $x_m$ is the $m$th sample in the LR training set $X$.

Defining $X_S$ and $Y_S$ as $X_S = \{x_m | x_m \in X, m \in S_t\}$ and $Y_S = \{y_m | y_m \in Y, m \in S_t\}$ respectively, we can rewrite the objective function Eq. (5) as the following matrix form:

$$O_{LR-LiSR} = \|PX_S - Y_S\|_F^2 + \alpha\|P\|^2_F.$$  

(8)

2.2. HR-LiSR

In above presented LR-LiSR, we obtain the support and target HR patch (constructed by projecting the LR input patch to the mapping matrix) separately and actually consider only one manifold (the LR patch manifold), thereby ignoring the geometrical information of the HR patch space, which is much more credible (i.e., the relationship between patches is true) than that of the LR one. In the following, we will further improve the LR-LiSR model from two aspects. Firstly, in contrast to estimate the support on the LR image patch manifold, we get the support from the HR image patch manifold. Secondly, we design an iterative schema to update the support and the mapping function. In this way, the estimated HR patch by the mapping function will benefit the generation of the support, while the newly obtained support will further improve the accuracy of the mapping function, leading to a perfect prediction for the HR image patch.

Based on above discussion, we give the objective function of the HR-LiSR model:

$$\hat{P}, \hat{S}_H = \arg \min_{P, S_H} \sum_{m \in S_H} (Px_m - y_m)^2 + \alpha\|P\|^2_F.$$  

(9)
Given the LR and HR training samples, X and Y, our goal is to obtain the support \( S_H \) and the mapping function \( P \) simultaneously. Here, \( S_H \) is defined on the HR training set given the estimated HR patch \( y_i \):

\[
S_H = \text{supp}(\text{dist}_H | k),
\]

where \( \text{dist}_H \in R^M \) is the measurement of squared distances between the estimated HR patch \( y_i \) and the HR training patches. Specifically,

\[
\text{dist}_{Hm} = \|y_i - y_m\|^2, \quad 1 \leq m \leq M,
\]

where \( y_m \) is the \( m \)th base in the HR training set \( Y \), \( \text{dist}_{Hm} \) refers to the smallest \( K \) entries of \( \text{dist}_H \), and \( \text{supp}(\text{dist}_H | k) \) is a set composing the indexes of these smallest \( K \) entries in \( \text{dist}_H \).

Defining \( X_S \) and \( Y_S \) as \( X_S = \{x_m | x_m \in X, m \in S_H \} \) and \( Y_S = \{y_m | y_m \in Y, m \in S_H \} \) respectively, we can rewrite the objective function Eq. (9) as the following matrix form:

\[
\langle \hat{P}, \hat{S}_H \rangle = \text{arg min}_{\hat{P}, \hat{S}_H} \|PX_S - Y_S\|^2 + \alpha \|P\|^2.
\]

(12)

The objective function of LR-LiSR is to minimize the mapping errors in the support \( S_H \). Given the support \( S_H \), Eq. (12) is quadratic in \( P \) and can be solved by many quadratic programming techniques. In this paper, we simply set the partial derivatives to zero to find the minimum.

Using matrix properties \( \text{tr}(AB) = \text{tr}(BA) \), \( \|A\|^2 = \text{tr}(A^T A) \), and \( \text{tr}(A) = \text{tr}(A^T) \), we have

\[
O_{HR-LiSR} = \|PX_S - Y_S\|^2 + \alpha \|P\|^2
\]

\[
= \text{tr}(PX_S - Y_S)(PX_S - Y_S)^T + \alpha \text{tr}(PP^T).
\]

(13)

In order to minimize the objective function Eq. (13), we would like to take the derivative of it with respect to \( P \) and set it to zero, i.e., we have the following equation:

\[
\frac{\partial O_{HR-LiSR}}{\partial P} = 2PX_SX_S^T - 2Y_SY_S^T + 2\alpha P = 0
\]

\[
\Rightarrow P = X_S(X_S^T)^{-1}(Y_S^T + \alpha I).
\]

(14)

2.3. Face image SR via LiSR

The face image SR problem can be formulated as the estimation of a HR image \( y_i \) from one input LR image \( x_i \), given a training set of HR images and their corresponding LR versions. We first decompose a complete face image into smaller patches according to the position, the patches are processed in raster-scan order in the image, from left to right and top to bottom. Then, we apply the proposed LR-LiSR or HR-LiSR model to each LR patch image to predict its HR version. Following [1] and [2], we enforce compatibility between adjacent patches (the values in the overlapped regions are simply averaged). Pseudocode of the LR-LiSR and HR-LiSR based face image SR approaches is given in Algorithm 1 and Algorithm 2 respectively.

Algorithm 1. Face image SR via LR-LiSR.

1: Input: The training sets of HR and LR face images, \( Y = \{y_m\}_{m=1}^M \) and \( X = \{x_m\}_{m=1}^M \), and an input LR face image \( x_i \). The parameter \( patch\_size \), \( overlap \), and \( K \).

2: Output: HR super-resolved face image \( y_i \).

3: Compute \( U \) and \( V \):

\[ U = \text{cell(imrow -- overlap)/(patch\_size -- overlap)} \]

\[ V = \text{cell(imcol -- overlap)/(patch\_size -- overlap)} \]

4: Divide each of the training images and the input LR image into \( N \) small patches according to the same location of face, respectively.

\[ \{y_m, x_m | y_m \in Y, x_m \in X, m \in S_H \} \]

where \( S_H \) is globally learned off-line in the image patch space and its measurement is overwritten in one time. In this subsection, we will specifically distinguish our work from an insightful perspective.

For LIT [28] and GSCR [34], the linear regression model is globally learned off-line in the image patch space and the sparse coding space respectively. In contrast, we locally construct the linear regression relationship for each observation patch on its support LR/HR patches, and the HR

Algorithm 2. Face image SR via HR-LiSR.

1: Input: The training sets of HR and LR face images, \( Y = \{y_m\}_{m=1}^M \) and \( X = \{x_m\}_{m=1}^M \), and an input LR face image \( x_i \). The parameter \( patch\_size \), \( overlap \), and \( K \).

2: Output: HR super-resolved face image \( y_i \).

3: Compute \( U \) and \( V \):

\[ U = \text{cell(immr -- overlap)/(patch\_size -- overlap)} \]

\[ V = \text{cell(imcol -- overlap)/(patch\_size -- overlap)} \]

4: Divide each of the training images and the input LR image into \( N \) small patches according to the same location of face, respectively.

\[ \{y_m, x_m | y_m \in Y, x_m \in X, m \in S_H \} \]

where \( S_H \) is globally learned off-line in the image patch space and its measurement is overwritten in one time. In this subsection, we will specifically distinguish our work from an insightful perspective.

For LIT [28] and GSCR [34], the linear regression model is globally learned off-line in the image patch space and the sparse coding space respectively. In contrast, we locally construct the linear regression relationship for each observation patch on its support LR/HR patches, and the HR
patch is obtained by the specific regression function. LLT and GSCR cannot be adjusted by the input LR patch image, while our method can assign different support to the input image, thus our method is much more flexible.

As for these kernel regression-based methods [26,27], they define the kernel function in the LR patch space given the LR input patch image. In contrast, in our proposed method (HR patch guided LiSR), we define the kernel (in terms of the support set) in the HR space (which is much more credible and discriminant than the LR one [31,39]), and iteratively update it through a objective function with two variables (refer to Eq. (9)). Additionally, in comparison with the above mentioned general SR method, we use the position prior of face image to construct the regression model for each position patch.

3. Experiments

In this section, we describe the details of extensive experiments performed to evaluate the usefulness of the proposed method for face image SR. The experiments are designed to answer the following four questions:

- How does the proposed approach compare against other state-of-the-art face image SR methods; does the proposed approach succeed in enhancing real-world images?
- Is it necessary to iterate the processes of searching K-NN in the HR patch space (generating the support) and learning the mapping function from support LR and HR patch pairs, and does the iterative process convergence?
- Does the locality constraint facilitate the proposed approach succeed in generating a better face image SR performance?
- How do the dictionary size and the global constraints (the refinement step) affect the performance of the SR methods?

### 3.1. Baselines and parameter settings

The SR experiments are conducted on two databases, the CAS-PEAL-R1 database [40] and the AR face database [41]. Brief descriptions of both datasets are provided later along with the details of the experiments. We compare the SR quality of our approach (both LR-LiSR and HR-LiSR) with the state-of-art algorithms subjectively and objectively. Some respective algorithms are selected to compare with our method which include Wang’s global face image SR method [22], three coding-based face image SR methods (Chang’s NE based method [1], Yang’s SC based method [2] and our previous proposed LcR based method [25]), and three regression-based methods (Kim’s KRR based general image SR method [27], Tang’s GSCR based general image SR method [34] and Huang’s LLT based face image SR method [28]). Note that many influential work have been proposed for general image SR, such as [33,21,42], we only give some representative method’s results in this section due to space limitation.

In order to obtain the best performance, we adjust the parameters for each comparative method. For Wang’s eigentransformation method, we let the variance accumulation contribution rate of PCA be 99%. For the sake of fair competition, we set the HR patch size to 12 × 12 in these patch-based methods and the overlap between neighbor patches is 4 pixels while the corresponding LR image patch size is set to 3 × 3 with overlap of 1 pixel. In the experiments, we find that the larger the overlap the better the final results. Obviously, the larger the overlap the longer the running time. Therefore, we must balance performance and efficiency. The number of neighbors in Chang’s NE is set to 50 for the CAS-PEAL-R1 database and is set to 100 for the AR face database. For Yang’s SC method and Kim’s KRR method, we have used the code available from the author’s website. For our previous proposed LcR method, the locality constraint parameter is set to 0.04 both in the CAS-PEAL-R1 database and the AR face database. Note that, to remove the blurring and ringing effects around strong edges because of the regression, we also add a correction step for Huang’s LLT method and Kim’s KRR method for refinement. More details about the correction step for refinement will be discussed in Section 3.4.5. For Tang’s GSCR method, we trained the coupled dictionaries with 1024 elements from 50,000 randomly sampled patches from the training face images. As for LR-LiSR, the size of nearest neighbor numbers K is set to 100 both in the CAS-PEAL-R1 database and the AR face database. As for HR-LiSR, it has only two free parameters, i.e., the size of nearest neighbor numbers K and the maximum iteration. In our experience, K is set to 100 and the maximum iteration is set to 3 both in the CAS-PEAL-R1 database and the AR face database, we will demonstrate the performance versus these two parameters in Sections 3.4.1 and 3.4.2, respectively.

### 3.2. Comparison with existing approaches on the CAS-PEAL-R1 database

In this subsection, we conduct experiments on the public face database, the CAS-PEAL-R1 database, with 30,871 images of 1040 subjects. We only use the neutral
expression and normal illumination face of each subject from the frontal subset for experiments. In all the 1040 frontal face images, we randomly select 1000 images for training and leave the other 40 images for testing. Therefore, all the test subjects are not present in the training images. All the images are aligned by two eyes and cropped to $112 \times 100$ pixels (some examples are shown in Fig. 4). The LR images are formed by smoothing (an averaging filter of size $4 \times 4$) and down-sampling (by a factor of 4) corresponding HR images, thus the size of LR face images are $28 \times 25$ pixels.

To evaluate the respective methods quantitatively, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [43], in terms of visual image quality assessment, are given in Table 1. We can see that our method outperforms the state-of-art methods in terms of both PSNR and SSIM. From the table, it seems that the best alternative is the patch-based SR methods (SC, LcR and our method) that incorporate position prior information. This is due to the strong representation ability of the patch image and is sufficient to reconstruct the global face image. The global method is sensitive to the input LR image, i.e., if the input LR image is dissimilar to the training samples, the super-resolved face will be incredible. As a general image SR, KRR is inapplicable to face image. GSCR learns the coupled dictionaries from training face images, and its performance is better than KRR but worse than those position patch based method. This indicates the importance of position prior to face image SR. While LLT is a patch-based method that also incorporates position prior information of face, its performance is obviously not as good as other position patch-based methods (SC, LcR and our method). This can be attributed to that the pre-trained transformation function may not be suitable for the various patches and it cannot be adjusted adaptively according to the given LR patch. The performance of our previous proposed LcR based method and our proposed method (both LR-LiSR and HR-LiSR) is significantly better than the other comparison methods, and it is mainly because of the introduction of locality constraint in them.

The HR images reconstructed by our proposed approach and the representative SR methods are shown in Fig. 5. It can be seen that Wang's global method gives

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang and Tang [22]</td>
<td>26.62</td>
<td>0.8254</td>
</tr>
<tr>
<td>NE [1]</td>
<td>27.91</td>
<td>0.8868</td>
</tr>
<tr>
<td>SC [2]</td>
<td>28.27</td>
<td>0.8968</td>
</tr>
<tr>
<td>LcR [25]</td>
<td>28.84</td>
<td>0.9083</td>
</tr>
<tr>
<td>KRR [27]</td>
<td>26.97</td>
<td>0.8597</td>
</tr>
<tr>
<td>GSCR [34]</td>
<td>28.18</td>
<td>0.8971</td>
</tr>
<tr>
<td>LLT [28]</td>
<td>27.64</td>
<td>0.8823</td>
</tr>
<tr>
<td>LR-LiSR</td>
<td>28.96</td>
<td>0.9093</td>
</tr>
<tr>
<td>HR-LiSR</td>
<td>29.30</td>
<td>0.9137</td>
</tr>
<tr>
<td>Improvement</td>
<td>0.46</td>
<td>0.0054</td>
</tr>
</tbody>
</table>

Fig. 5. Comparison of results based on different methods on the CAS-PEAL-R1 database: (a) input LR faces; (b) Wang and Tang [22]; (c) NE [1]; (d) SC [2]; (e) LcR [25]; (f) KRR [27]; (g) GSCR [34]; (h) LLT [28]; (i) LR-LiSR; (j) HR-LiSR; (k) original HR faces. (Note that the effect is more pronounced if the figure of the electronic version is zoomed.)
the dirty and jagged results, such as the eyes, mouths and noses, and cannot effectively reconstruct the fine individual details of the novel testing faces for its holistic. The result in Fig. 5(c) is derived from the NE method. This method is the most representative manifold learning based method, differing essentially from our method in terms of geometrical structure information being used to preserve the similar local structure. In NE, the local geometrical structure of the input LR patch is utilized to guide the reconstruction of target HR patch. In contrast, our method estimates the local geometrical structure upon the local structures of the estimated HR patch in the previous step in the HR patch manifold. In other words, our method borrows extra information from the HR patch manifold given the estimated HR patch, and not by using the LR observations only, as in NE. This also explains why NE produces blurry results and incorrect facial details.

The images in Fig. 5(d) and (e) are generated by two patch-based face image SR methods: SC and LcR. SC adopts a sparsity constraint to super-resolved the HR facial images, while LcR introduces locality constraints to regularize the model to obtain a stable solution. As can be seen from the results, plausible facial structures can be well inferred in the resulting HR images. Nevertheless, as these methods are based on the assumption that the LR and HR patch manifolds share the similar local geometrical structure, which is not always the case in practical terms, they cannot effectively reconstruct the fine individual details. KRR is a general image SR method, and the learned prior model of a generic image would not apply to face images. Similarly, the gradient profile prior [42], which is learned from the general and does not consider the specific properties of human faces image, may not be suitable for face image also. In addition, as a kernel regression method, KRR defines the kernel function in the LR image space, which is much more credible and discriminant than the LR one [31,39]. Therefore, the super-resolved faces of KRR are smooth and lack of high frequency information. GSCR randomly sampled the LR and HR patch pairs from the training set, neglecting the position prior, thus the super-resolved face images are smooth faces and are with artifacts (Fig. 5(g)). Fig. 5(h) is the results of LLT, which applies a global patch regression model to establish the relationship between the LR and HR patches in the training set, and then utilizes the established model to predict the target HR patch of the input LR patch. When compared with our results, results of LLT are very smooth and lack of fine facial details. This is mainly because that the regression model of LLT is learned by the whole training samples without considering the observation LR patch. In other words, it cannot be adaptively adjusted given the observation LR patch. Considering this deficiency, we model the regression relationship between the support LR/HR patch pairs instead of the whole training samples. In addition, to better explore the local geometrical structure, we further proposed an iterative mapping strategy in the adaptively adjusted support domain. As shown in Fig. 5(i) and (j), the super-resolved faces are more reasonable and have much more fine facial details and clear face contours.

3.3. Comparison with existing approaches on the AR face database

In this subsection, we conduct experiments on the AR face database to testify the robustness of the proposed method to occlusions, i.e., wear sunglasses or occluded with scarves, which often exist in actual application. This experiment is performed on a subset of the AR face database. It has 600 frontal facial images corresponding to 100 subjects (containing 50 males and 50 females), i.e., every person has six frontal facial images, two of which wear sunglasses and two of which are occluded with scarf, as shown in Fig. 6. Similarly, in all the 600 frontal face images corresponding to 100 subjects, we randomly select 90 subjects (540 images) for training and leave the other 10 subjects (60 images) for testing. Therefore, the test subjects are also not present in the training samples. All the images are aligned by two eyes and cropped to $112 \times 88$ pixels. The LR images are formed by the same way as in the CAS-PEAL-R1 database, thus the size of LR face images is $28 \times 22$ pixels.

We show the PSNR and SSIM of all 60 face images in Table 2. We can get the same conclusion that our method is superior or competitive to the other SR approaches in terms of both PSNR and SSIM. In addition, we also give the super-resolved results in Fig. 7. It shows that our method.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang and Tang [22]</td>
<td>25.13</td>
<td>0.7030</td>
</tr>
<tr>
<td>NE [1]</td>
<td>28.16</td>
<td>0.8572</td>
</tr>
<tr>
<td>SC [2]</td>
<td>28.80</td>
<td>0.8603</td>
</tr>
<tr>
<td>LcR [25]</td>
<td>29.35</td>
<td>0.8738</td>
</tr>
<tr>
<td>KRR [27]</td>
<td>25.72</td>
<td>0.8213</td>
</tr>
<tr>
<td>GSCR [34]</td>
<td>28.10</td>
<td>0.8424</td>
</tr>
<tr>
<td>LLT [28]</td>
<td>28.10</td>
<td>0.8424</td>
</tr>
<tr>
<td>LR-LiSR</td>
<td>29.16</td>
<td>0.8695</td>
</tr>
<tr>
<td>HR-LiSR</td>
<td>29.43</td>
<td>0.8745</td>
</tr>
<tr>
<td>Improvement</td>
<td>0.08</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

![Fig. 6. Some samples in the AR face database.](image-url)
gives the best results when occlusions occur. Our method can recover the detailed face features of noses, mouths, and face contours so well that it is almost free of ringing effects. Wang’s results have a ghost effect on the face contours and distortion around features especial around eyes and noses. NE may construct some details that do not exist in the original HR faces or lose some important information, e.g., the glass of the first row. KRR and LLT get the blur face images, in addition, there are some ringing effects around face counters.

![Comparison of results based on different methods on the AR face database](image)

Fig. 7. Comparison of results based on different methods on the AR face database: (a) input LR faces; (b) Wang and Tang [22]; (c) NE [1]; (d) SC [2]; (e) LcR [25]; (f) KRR [27]; (g) GSCR [34]; (h) LLT [28]; (i) LR-LiSR; (j) HR-LiSR; (k) original HR faces.

![Graph showing PSNR and SSIM](image)

Fig. 8. The influence of the number of different iteration.
3.4. Influence of parameters on reconstruction results

We next consider the influence of various parameters on the super-resolved face images of the proposed method. Here we only show the experimental results on the CAS-PEAL-R1 database for limited space, and we can obtain the similar conclusion on the AR face database.

3.4.1. Influence of iteration number

As shown in Fig. 8, we plot the average PSNR and SSIM values of all the test face images according to the number of iteration, and find that (1) as the iteration number increases, the gain of the proposed HR-LiSR method becomes larger, which implies that HR patch manifold structure guided K-NN searching is vitally important for the regression model. That is to say, the accurate HR patch estimation will further improve the accuracy of support information and promote the prediction performance of the learned regression model; (2) the proposed HR-LiSR method can be fast-converging with a small number of iterations, i.e., the iteration number is set to 3 for all experiments, which implies a potential application in practice.

3.4.2. Influence of nearest neighbors K

Fig. 9(a) is the plot of average PSNR of all 40 test face images according to different values of K. To see what happens when K change between 1 and 100, we plot the average PSNR in Fig. 9(b), and the curves present an “up-down-up” behavior. In addition to the individual curves, we also tabulate the average PSNR values with typical K in Table 3. We can see that this parameter has big implications for the performance of the proposed method. The fall around 9 can be explained by that the regression model is “too fitted” on the training data (as we use $3 \times 3$ LR patches, the regression model will over-fit the training samples and this phenomena has been observed by [2] and [44]) and thus generates undesired bad HR reconstructions. For the proposed HR-LiSR method, we set the number of nearest neighbors K to 100.

3.4.3. Influence of support

We test that how the support impacts the final face image SR performance. In particular, we generate the support by four strategies: the LR patch guided K-NN support generation method (the LR-LiSR model), the estimated HR patch guided K-NN support generation method (the HR-LiSR model), the random support generation method, the estimated HR patch

---

**Table 3**

PSNR(dB) performance with different nearest neighbors K.

<table>
<thead>
<tr>
<th>K</th>
<th>3</th>
<th>6</th>
<th>9</th>
<th>12</th>
<th>15</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.8243</td>
<td>0.8111</td>
<td>0.3036</td>
<td>0.8257</td>
<td>0.8700</td>
<td>0.8903</td>
<td>0.9099</td>
<td>0.9137</td>
<td>0.9127</td>
<td>0.9103</td>
<td>0.9046</td>
</tr>
</tbody>
</table>

---

**Fig. 9.** The average PSNR and SSIM curves with respect to the number of nearest neighbors K. The fall around 9 can be explained by that the least squares solution of neighbor embedding is “too fitted” on the LR data.
guided sparse support generation method. We list the average PSNR and SSIM performance of these four different support generation strategies according to different values of $K$ in Fig. 10. We also give the best performance of these four strategies with optimal $K$ as well as the performance when $K=1000$ in the captions of Fig. 10.

The performance of the proposed HR patch guided support generation method can achieve the best performance when $K$ is set to a proper value, i.e., $K=100$. As can be seen, the average PSNR and SSIM improvements of these four methods are 29.30, 28.95, 28.54, 27.98 and 0.9137, 0.9096, 0.9046, 0.8924, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

![Fig. 10. The performance of four different support estimation strategies. The best average PSNR and SSIM of these four methods are 29.30, 28.95, 28.54, 27.98 and 0.9137, 0.9096, 0.9046, 0.8924, respectively. When $K=1000$, The average PSNR and SSIM of these four methods are 28.55, 28.16, 28.55, 27.98 and 0.9046, 0.8979, 0.9046, 0.8924, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)](image)

Table 4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>26.62</td>
<td>27.91</td>
<td>28.27</td>
<td>28.84</td>
<td>27.64</td>
<td>28.96</td>
<td>29.30</td>
</tr>
<tr>
<td>500</td>
<td>25.27</td>
<td>27.71</td>
<td>28.15</td>
<td>28.72</td>
<td>27.62</td>
<td>28.65</td>
<td>28.95</td>
</tr>
<tr>
<td>200</td>
<td>23.38</td>
<td>27.44</td>
<td>27.88</td>
<td>28.39</td>
<td>27.50</td>
<td>28.20</td>
<td>28.43</td>
</tr>
<tr>
<td>100</td>
<td>22.05</td>
<td>27.23</td>
<td>27.65</td>
<td>28.05</td>
<td>27.41</td>
<td>27.85</td>
<td>28.00</td>
</tr>
</tbody>
</table>

Increasing the value of $K$. This is primarily because a large $K$ (select almost all the bases as the support) implies that the support has less influence on the recovery. When $K=1000$, the only difference between the HR patch guided $K$-NN support and the LR patch guided $K$-NN support is that the former has an additional iteration step. And we can see that this additional iteration step gain 0.39 dB and 0.0067 improvements in terms of PSNR and SSIM respectively. This certifies the effective of our iterative framework again.

3.4.4. Influence of training set size

The above experimental results show that the sparsity prior in Yang’s SC model and the locality prior in our previous LcR model and the proposed LR-LiSR and HR-LiSR models are very effective in regularizing the ill-posed SR problem. In those methods, we fixed the training set size to be 1000. Intuitively, the larger training set should possess more expressive power and thus may yield more accurate approximation, while increase the running time. In this subsection, we evaluate the effect of dictionary size on face image SR. We randomly select three training subsets of size 100, 200, and 500, and apply them to the same input image. Table 4 tabulates the average PSNR of all 40 test images of different methods (Note that the KRR and GSCR are independent of the training set, we would not consider them here.) With the decreasing of the training set size,
the average performance of all methods also decreases. The fastest decline algorithm is Wang’s global face image SR. This is mainly because Wang’s method is based on a statistical mode, which could not reveal the distribution of data when the training samples are not sufficient (the feature dimension is very large, e.g., $28^2/25 = 700$). From the results, we also observe that with the increase of training set size, the performance gain of our proposed LR-LiSR and HR-LiSR methods over the other three coding-based methods tends contractible. Unsurprisingly, when the training set is small, it is much more difficult for LiSR to obtain the reasonable support, thus the learned regression model may not be suitable for the LR observation.

### 3.4.5. Influence of global constraints

In order to refine the super-resolved results, many post-processing approaches have been proposed [2,32]. They enforce the global reconstruction constraint to refine the patch-based model, ensuring the recovered HR image to be consistent with its LR observation. In our experiments, we observe that our proposed regression model is very effective and contributes the most, while the global constraint in post-processing produces a very small gain. Unsurprisingly, when the training set is small, it is much more difficult for LiSR to obtain the reasonable support, thus the learned regression model may not be suitable for the LR observation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patch-based</td>
<td>24.49</td>
<td>27.91</td>
<td>28.27</td>
<td>28.84</td>
<td>25.31</td>
<td>28.18</td>
<td>27.06</td>
<td>28.96</td>
<td>29.30</td>
</tr>
<tr>
<td>Plus global</td>
<td>26.17</td>
<td>28.28</td>
<td>28.29</td>
<td>28.86</td>
<td>26.91</td>
<td>28.21</td>
<td>27.64</td>
<td>28.98</td>
<td>29.32</td>
</tr>
<tr>
<td>Improvement</td>
<td>1.68</td>
<td>0.37</td>
<td>0.02</td>
<td>0.03</td>
<td>1.60</td>
<td>0.03</td>
<td>0.58</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 5: Global constraint in the post-processing further refines the PSNR(dB) performance from patch-based model in the first step.

The input LR face images of all the above experiments are formed by simply smoothing and down-sampling HR images, which cannot represent the real complex spatial feature relationship between the HR image and the degraded LR one as they do not correspond to the images captured by a real camera [45]. In an actual condition, it is too difficult for us to simulate the image degradation process. Therefore, in order to further show the efficiency of the proposed framework in real-world situations, we perform experiments on four face images captured by real surveillance imaging condition\(^1\) and seven face images from the CMU + MIT face database [46]. We show the original pictures, the extracted LR face images and the super-resolved HR face images using our proposed HR-LiSR method in Fig. 12. We can see that our proposed method can produce reasonable results for different degradation processes even though the test LR images are different from the training examples and contaminated with heavy noise.

### 3.5. SR with real-world face

The input LR face images of all the above experiments are formed by simply smoothing and down-sampling HR images, which cannot represent the real complex spatial feature relationship between the HR image and the degraded LR one as they do not correspond to the images captured by a real camera [45]. In an actual condition, it is too difficult for us to simulate the image degradation process. Therefore, in order to further show the efficiency of the proposed framework in real-world situations, we perform experiments on four face images captured by real surveillance imaging condition\(^1\) and seven face images from the CMU + MIT face database [46]. We show the original pictures, the extracted LR face images and the super-resolved HR face images using our proposed HR-LiSR method in Fig. 12. We can see that our proposed method can produce reasonable results for different degradation processes even though the test LR images are different from the training examples and contaminated with heavy noise.

\(^1\) The low-quality picture with a CIF-size (352×288 pixels) is taken by a surveillance camera in the condition of underexposure and the interested object is at a distance. The input LR face images are obtained by converting the extracted faces to grayscale and simply adjusting the levels.
4. Discussion

4.1. Global vs. local

The global face based SR method, e.g., Wang’s eigentransformation method [22], can capture the global structure of face. However, it will also result in low reconstruction precision and unsatisfactory results around the face contours and margin of the mouth. This requires an additional step to compensate for residual errors and preserve characteristics [17,47,32]. Since the local patch-based model has higher reconstruction precision than the global model, it can be successfully applied in preserving the global structure.

In addition, the locality is very important for local patch representation. By the introduction of locality constraint, our previously proposed method, i.e., LcR [25], and the proposed LR-LiSR and HR-LISR methods all outperform
other local patch-based methods, which alternatively represents the patch sparsely and collaboratively.

4.2. Embedding LR patch manifold structure vs. embedding HR patch manifold structure

Due to the “one-to-many” mapping between the LR image and the HR image, the neighborhood relationship of the LR image patch manifold could no longer reflect the inherent geometrical structure. Thus, the assumption that the LR and HR image patch manifolds share the same local geometrical structure is inappropriate in practical applications. Traditional manifold embedding based SR methods [1,2] reconstruct the HR image by embedding the local structure of the LR image manifold into the reconstructed HR image manifold. Obviously, based on the above analysis, it is unreasonable. The proper way should be embedding the local structure of the HR image patch manifold into the reconstructed HR one. However, we cannot know the desired HR image in advance, thus the local structure of the HR image manifold is unknown. Therefore, we seek for an iteration approach to respectively and iteratively obtain the HR image and its local structure, which will be discussed in the following subsection.

4.3. Package vs. iteration

Traditional approaches including our previous proposed LcR method all aim at designing a package strategy to infer a HR image. The common idea of the traditional methods is to preserve the manifold structure of the LR patch manifold for that of the corresponding LR patch manifold. Actually they consider only one manifold (LR patch manifold), while ignoring the geometrical information of the HR patch manifold, which is much more credible and discriminant [31,39]. From our experiments, we can learn that the iterative optimization is very important for the LR and HR patch manifold preservation. In other words, the previous estimated HR patch can be applied to the exploration of the manifold structure, which in turn will promote the accurate inference of HR patch.

5. Conclusions and future work

In this paper, we proposed a novel locality-induced Support Regression (LiSR) framework for single frontal face image SR. Under LiSR, given a low-resolution (LR) input patch, its support and a mapping function will be learned. And then, the target HR image can be predicted by the learned mapping function on the support LR/HR patch pairs. In order to obtain the support, we design two strategies by incorporating the locality constraint of manifold structure, i.e., LR patch guided support generation and HR patch guided support. These will lead to two varieties of LiSR, LR-LiSR and HR-LiSR. In HR-LiSR, as we could not obtain the target HR patch beforehand, we further designed an iterative algorithm to learning the mapping from LR to HR patches and generating the support. The proposed LiSR model is adapted to single frontal face image SR and shows very competitive performance with state-of-the-arts.

In this paper we mainly focus on super-resolving the frontal face. However, in the wild-world, the captured LR face may contain pose variety. The straightforward approach is to pre-construct the non-frontal LR training set (with the same pose as the captured LR face), and learn the relationship between the non-frontal LR training set and the corresponding HR one to infer the target HR face image. But in practice, we cannot pre-construct training samples with arbitrary poses (according to the input faces). In the future, we will introduce some more advanced image representations and preprocessing techniques to further improve the results.

As pointed in [48], the reconstruction constraint, namely that the blurred and downsampled HR output should approximately equal the LR input image, has been either ignored or applied with default fixed blur models in most previous methods as well as our proposed method. Therefore, incorporating the reconstruction constraint to further improve the SR results will be another future work.

Acknowledgments

The research in this paper use the CAS-PEAL-R1 face database collected under the sponsor of the Chinese National Hi-ech Program and ISVISION Tech. Co. Ltd. The research was supported by the major national science and technology special projects (2010ZX03004–003–03), the National Key Technologies R&D Program (2013AA014602), the National Natural Science Foundation of China (60970160, 61070080, 61003184, 61172173, 61303114, 6170023, and 61303154), the China Postdoctoral Science Foundation funded project (2013M530350), Specialized Research Fund for the Doctoral Program of Higher Education (SRFPD) (201301120024), the President Fund of UCAS, the Open Project Program of the National Laboratory of Pattern Recognition.

References
