Information Sciences 367–368 (2016) 354–372

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation

Junjun Jiang[a,*], Chen Chen[b], Kebin Huang[c], Zhihua Cai[a], Ruimin Hu[c]

[a] School of Computer Science, China University of Geosciences, Wuhan 430074, China
[b] Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA
[c] National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 15 January 2016
Revised 16 May 2016
Accepted 23 May 2016
Available online 7 June 2016

Keywords:
Super-resolution (SR)
Face hallucination
Neighbor embedding
Tikhonov regularization
Low-resolution

ABSTRACT

In human-machine interaction, human face is one of the core factors. However, due to the limitations of imaging conditions and low-cost imaging sensors, the captured faces are often low-resolution (LR). This will seriously degrade the performance of face detection, expression analysis, and face recognition, which are the basic problems in human-machine interaction applications. Face super-resolution (SR) is the technology of inducing a high-resolution (HR) face from the observed LR one. It has been a hot topic of wide concern recently. In this paper, we present a novel face SR method based on Tikhonov regularized neighbor representation (TRNR). It can overcome the technological bottlenecks (e.g., instable solutionand noise sensitive) of the patch representationscheme in traditional neighbor embedding based image SR methods. Specifically, we introduce the Tikhonov regularization term to regularize the representation of the observation LR patches, leading to a unique and stable solution for the least squares problem. Furthermore, we show a connection of the proposed model to the neighbor embedding model, least squares representation,sparse representation, and locality-constrained representation. Extensive experiments on face SR are carried out to validate the generality, effectiveness, and robustness of the proposed algorithm. Experimental results on the public FEI face database and surveillance images show that the proposed method achieves better performance in terms of reconstruction error and visual quality than existing state-of-the-art methods.

1. Introduction

In real-world scenarios, low-resolution (LR) images are generally captured in many electronic imaging applications, such as video surveillance, consumer photographs, satellite imaging, magnetic resonance imaging and video standard conversion. The resolution of captured images is limited by image charge-coupled device (CCD) and complementary metal-oxide semiconductor (CMOS) image sensors, hardware storage and other constraints in electronic imaging systems. However, high-resolution (HR) images or videos are usually desired and often required for the following face synthesis and analysis. Image super-resolution (SR) reconstruction, as an effective way to solve this problem, aims to reconstruct HR images from the observed LR ones. It can increase high-frequency components and remove the undesirable effects, e.g., resolution degradation, blur and noise (see Fig. 1). In this paper, we mainly focus on the face SR (sometimes also called face hallucination) problem.

* Corresponding author. Tel.: +8615972034993; fax: +8602767883716.
E-mail address: junjun0595@163.com (J. Jiang).

http://dx.doi.org/10.1016/j.ins.2016.05.032
0020-0255/

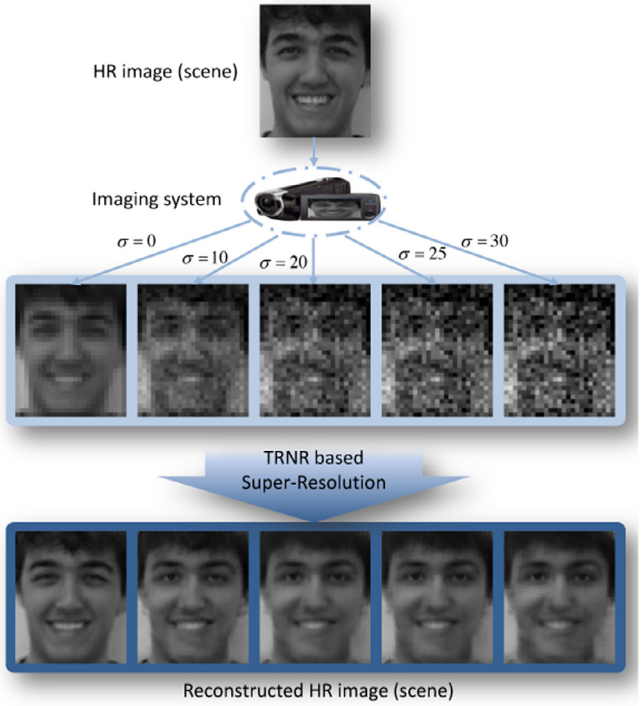


Fig. 1. Face SR reconstruction results of our proposed TRNR based method under different noise levels.

In the following, we will review some representative works in this field. For a more complete and comprehensive survey of image SR, please refer to [36].

1.1. Prior works

SR reconstruction is of great importance for various applications, and many algorithms have been developed in recent years. Generally speaking, they can be categorized based on their tasks, i.e., generic image SR [9,13,21,23,40,44,46,48] and domain-specific image SR focusing on specific classes of images such as faces [26,36] and text images [8].

In 2000, Baker and Kanade [2] developed a learning-based method named "face hallucination", which was the pioneering work on face SR. Given an input LR image, this approach infers the high frequency components from a parent structure with the assistance of training samples. Abandoning the Markov random field (MRF) assumption as in [9], this method is established based on training images using Gaussian, Laplacian and feature pyramids. Since the introduction of this work, a number of different methods and models have been developed for estimating the image information lost in the down-sampling process. These methods differ in ways of modeling the HR image. A successful face SR algorithm has to meet the global constraint (i.e., the results should have common human face characteristics) and the local constraint (i.e., the results should have specific characteristics of a particular face image). To fulfill these two constraints, Liu et al. [26] described a two-step approach integrating a principal component analysis (PCA) based global parametric model and a patch-based non-parametric Markov network. In the first step, the relationship between the HR face images and their smoothed, down-sampled LR ones is learned to obtain the global HR image. In the second step, the residual between an original HR image and its reconstructed one is compensated to learn the residual image. Both the work in [2] and [26] have shown that knowing an image containing makes it possible for an SR system to perform much better because the system can leverage regularities in face appearance to recover more details than creating from a general image model.

The above methods use probabilistic models and are based on an explicit resolution reduction function which sometimes is difficult to acquire in practice. Instead of using a probabilistic model, Wang et al. [37] proposed a face SR approach using eigentransformation algorithm, which can well capture the structural similarity of face images. This method treats the image SR problem as the transformation between LR images and their HR counterparts. An input LR image can be expressed as a linear combination of the LR images in the training set by using PCA. Then, the corresponding target HR image can be reconstructed by the same combination of the corresponding HR images in the training set. Following [37], many PCA based face SR methods have been proposed recently [4,30]. However, they are highly dependent on the training set since the core concept is to use a linear combination of the HR images in the training set to reconstruct the target HR face image. To achieve a good performance, these algorithms require that the training set size is very large, so that the test images have a great chance to be similar to the images in the training set.

Due to the poor representative ability, decomposing a complete face image into small face patches has been attracted much attention recently [15,16,19,25,32,39,40]. For example, Chang et al. [5] introduced locally linear embedding [31] and proposed a neighbor embedding (NE) based image SR method. It estimates HR patches by linearly combining K candidate HR patches in the training dataset. The algorithm assumes that an LR image patch and its HR counterparts (or feature representations) are locally isometric (i.e., share the same neighborhoods with the same representation coefficients). In [45], Zhang and Cham further extended the neighbor embedding algorithm to the discrete cosine transform (DCT) space. Huang et al. [16] applied the canonical correlation analysis (CCA) to perform a nonlinear mapping between HR and LR face images in a coherent space. In [1], Ae et al. extended CCA to 2D CCA. Hu et al. [15] proposed to use the local pixel structures and introduced a local structure learning based face SR method. To further explore the relationship between local pixels, Shi et al. [32] proposed to combine the global reconstruction model, the local sparsity model and the pixel correlation model into a unified regularization framework. Jiang et al. [20] claimed that there is a common hidden manifold between the LR and HR manifold spaces and proposed a manifold alignment based face SR method.

To incorporate more prior information about human face, which is a highly structured object, Ma et al. [29] assumed that similar textures are shared by the patches located at the same position of face images, and proposed a position-patch based face SR method. In their proposed method, HR patches are reconstructed as a linear combination of training position-patches using the same representation coefficients generated on the LR training samples by least squares representation (LSR). When the number of training position-patches is much larger than the dimension of patches, which is common in image SR, the least squares solution will not be unique. This will lead to unstable face SR results. Inspired by some sparse representation technologies [22,24,41], Yang et al. [40] regarded the training position-patches as an over-complete dictionary and proposed local patch sparse coding (SC) based face SR frameworks. Wang et al. [39] further proposed a weighted adaptive sparse regularization (WASR) method to super-resolve face images. In [32], kernel-based face hallucination dual regularization priors was proposed. The LR patches are mapped into an HR feature space. By taking into account the nonlinear structures, it is effective to estimate the combination coefficients in kernel feature space. By decomposing a face image into small face patches and incorporating the sparsity prior, it can well characterize the face structure and produce good HR results. However, when the input LR face is corrupted by noise, the sparse coding coefficients may be very unstable. Most recently, our previous work [19] further improved these position-patch based face SR methods by incorporating the local geometrical constraint of manifold rather than the sparsity constraint into the patch reconstruction objective function as in [40]. This locality-constrained representation (LcR) method can achieve sparsity and locality simultaneously. Due to the locality constraint, it can extract the maximum amount of facial information from the LR face images and is robust to noise. In [18], we further proposed a multilayer locality-constrained iterative neighbor embedding (MLINE) method to improve the reconstruction results of LcR by an iteration neighbor embedding strategy. However, when MLINE encounters noise in the iteration, it will amplify the noise resulting in poor SR performance. As far as we know, the LcR based face SR method [19] obtained the best performance reported in the literature, especially with noise input. It collaboratively represents the input LR patch image using all the training patches, therefore exhibiting strong representation power. However, when the input patch is contaminated by noise, LcR method may "*synthesize*" the noisy as well.

1.2. Motivation and contributions

Inspired by searching and weighting strategies in nonlocal means denoising method [3], which can remove most of the input noise and maintain primitive facial features, in this paper we propose a novel noise robust position-patch based representation method called Tikhonov regularized neighbor representation (TRNR). It uses a threshold (neighborhoods K) to search the best matching neighbor patches from the training samples and discard patches that are far from the input patch. Then, the target patch is reconstructed by weighting the obtained neighbor patches through locality-constrained neighbor embedding. Particularly, we introduce the Tikhonov regularization [12] to the least squares problem in neighbor embedding to give preference to a particular solution with desirable properties. Specifically, samples that are similar to the observation will have large freedom (thus have relative large reconstruction weights), while samples that are far away from the observation will be penalized (thus have very small reconstruction weights). Therefore, by this weighted Tikhonov regularization, the patch representation problem will achieve a stable solution. Experimental results demonstrate that our proposed TRNR method can preserve face structure and suppress most of the noise, and it is superior to the state-of-the-art methods in terms of reconstruction error and visual quality. The main contributions of our work can be summarized as follows:

- We propose a novel position-patch face SR method based on TRNR. It can be viewed as the generalization of several existing position-patch based methods, e.g., Chang et al.'s NE method [5], Ma et al.'s LSR method [29], Yang et al.'s SC method [40], and our previously proposed LcR method [19].
- By discarding the dissimilar training patches and imposing the Tikhonov regularization to the selected similar patches, TRNR can simultaneously remove noise and recover HR faces. Furthermore, the solution of TRNR model achieves both sparsity and locality, which have been proved to be two critical priors in exploring the face data structure.
- In addition to conduct experiments on the standard FEI face database [33], we also construct a test room with adjustable light sources to obtain surveillance face images at different distances and lighting conditions, and validate the effectiveness of our proposed method on these captured surveillance face images.

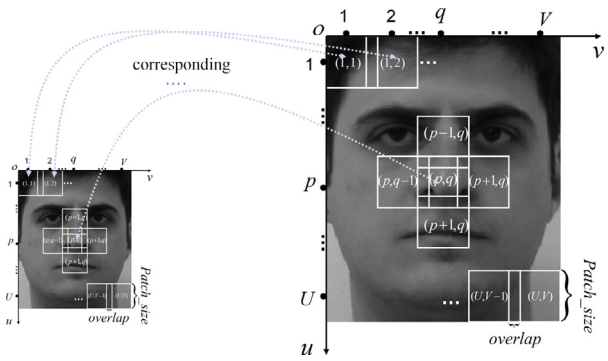


Fig. 2. Illustration of dividing a face into $M = UV$ patches. The term (p, q) indicates the coordinate of a patch in the patch coordinate system $O - uv$. *patch_size* and *overlap* denote the patch size of one square patch and the overlap pixels between neighbor patches respectively.

The remainder of this paper is organized as follows. Section 2 summarizes the notations used in this paper. In Section 3, we show the framework of position-patch based face SR method and briefly review four representative works. Section 4 provides the detailed descriptions of the proposed TRNR as well as the optimization of TRNR. We show the connections of the proposed method to existing position-patch based methods in Section 5. The experimental results and analysis on a standard face recognition database and some surveillance face images are given in Section 6. Section 7 summarizes the advantages and limitations of this work. The conclusion and future work are drawn in Section 8.

2. Notations

Given an LR observation image I_t^L (subscript "t" distinguishes test images from training images), our goal is to construct its HR version I_t^H by learning the relationship between the LR and HR training sets, $I^L = \{I_1^L, I_2^L, \cdots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \cdots, I_N^H\}$, where N is the number of samples in the training set. For a local patch based method, the input LR image is divided into M patches, $\{x_t(p, q)|1 \le p \le U, 1 \le q \le V\}$, according to the predefined *patch_size* and *overlap* pixels. As shown in Fig. 2, $x_t(p, q)$ denotes a small patch at the position (p, q) of the input LR image in the patch coordinate system, U represents the patch number in each column, V represents the patch number in each row, and $M = UV$. U and V can be calculated in the same way. The LR and HR training face image pairs are divided into M patches respectively, $\{x_i(p, q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$ and $\{y_i(p, q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$. $x_i(p, q)$ denotes a small patch at the position (p, q) of the i-th training sample in the LR training set, while $y_i(p, q)$ denotes a small patch at the position (p, q) of the i-th training sample in the HR training set. For the LR observation patch $x_t(p, q)$ at the position (p, q), its HR patch $y_t(p, q)$ is estimated using the LR and HR training image patch pairs at the same patch position. When it does not lead to a misunderstanding, we drop the term (p, q) for convenient from now on.

3. Position-patch based face SR

For the position-patch based methods, the common idea is firstly to exploit the relationship in the LR patch space formed by all the LR training image patches, and then apply this relationship to the HR patch space formed by all the HR training image patches. The position-patch based methods first represent a given LR patch x_t using the LR training image patch set $\{x_i\}_{i=1}^N$, and then utilize the representation coefficients w (i.e., the outcome of different representation methods) to estimate each corresponding (unknown) HR patch y_t by replacing the LR training image patch set $\{x_i\}_{i=1}^N$ with its HR counterpart $\{y_i\}_{i=1}^N$, i.e., $y_t = Yw$. To handle the compatibility problem between adjacent patches, simple averaging in the overlapping regions is performed. The key issue of these position-patch based methods is to obtain the optimal representation coefficients w, and this section reviews some existing representation schemes.

3.1. Neighbor embedding (NE)

In NE based image SR algorithm [5], it assumes that the LR and HR patch pairs share the same local geometrical structure, which is characterized by how an LR patch can be reconstructed by its neighbors in the LR patch space.[1] For each patch x_t in the LR image, the NE based method computes the reconstruction weights with its K nearest neighbors (K-NN) by minimizing the local reconstruction error. Then the HR embedding (patch) can be estimated with the corresponding HR patch neighbors

[1] It is worth noting that the original NE method was proposed for general images without any consideration of face priors, e.g., the position-patch prior. Therefore, in this paper we introduce the position-patch prior to the NE method for a fair comparison.

and the same reconstruction weights. Particularly, for each LR patch x_t in the input LR image, the optimal reconstruction weights are obtained by minimizing the local reconstruction error:

$$J_{\mathrm{NE}}(w) = \left\| x_t - \sum_{k \in C(x_t)} w_k x_k \right\|_2^2 \quad \text{s.t. } \sum_{i=1}^{N} w_i = 1, \tag{1}$$

where $C(x_t)$ is the index set of the K-NN of x_t in the LR training set.

3.2. Least squares representation (LSR)

Based on the manifold assumption, Ma et al. [29] presented an LSR based face SR approach. Given an LR observation patch x_t in the input LR face image, LSR uses patches from all training samples at the same position to represent x_t:

$$x_t = \sum_{i=1}^{N} w_i x_i + e, \tag{2}$$

where e is the reconstruction error. The optimal weights can be solved by the following constrained least square fitting problem:

$$J_{\mathrm{LSR}} = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2, \quad \text{s.t. } \sum_{i=1}^{N} w_i = 1, \tag{3}$$

where $w = [w_1, w_2, \ldots, w_N]^T$ is the optimal weight vector for the LR observation patch x_t. Above least square estimation can produce biased solutions when the dimension of the patch is smaller than the size of the training set.

3.3. Sparse coding (SC)

To solve the biased problem of (3), a possible way is to impose a regularization term. For example, [40] introduced the sparse representation theory and used a small subset of patches to represent LR observation patch x_t instead of performing collaboratively over all the training samples. These sparsity-based regularization models have shown promising results in image SR. Mathematically, these models assume that an observation signal x_t can be represented as $x_t \approx Xw$, where X is a over-complete dictionary , and most of the coefficients in w are close to zero due to the sparsity constraint. The sparsest w can be obtained by solving an optimization problem, $\min ||w||_0$, s.t. $||x_t - \sum_{i=1}^{N} w_i x_i||_2^2 \le \varepsilon$, where $\varepsilon \ge 0$ is an error tolerance, and $||\bullet||_0$ denotes the ℓ_0 -norm (which counts the number of nonzero entries). However, the problem is NP-hard and known to be intractable. To obtain a computationally practical and provably correct solution, it is often relaxed to the convex ℓ_1-norm optimization problem:

$$\min ||w||_1, \quad \text{s.t. } \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2 \le \varepsilon, \tag{4}$$

or the penalty version:

$$J_{\mathrm{SC}} = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2 + \lambda ||w||_1, \tag{5}$$

where $||\bullet||_1$ denotes the ℓ_1-norm and λ is a sparsity regularization parameter. This sparsity constraint not only ensures that the under-determined equation has an exact solution but also allows the learned representation for each patch to capture salient properties. Eq. (5) is known as LASSO or basis pursuit. There are many existing methods can be used to solve the problem, such as LASSO ℓ_1-norm regularization [22] and feature sign search algorithm [34].

3.4. Locality-constrained representation (LcR)

From formulation (5), we can see that SC based method reconstructs the input LR patch via a linear combination of a small number of training samples. Geometrically speaking, it is more reasonable to represent x_t by a linear combination of only its nearest neighbors in the training set to capture the local geometrical information of data. The recent theoretical results in machine learning have shown that the learning performance can be significantly enhanced if the local geometrical structure is exploited [31]. Sparse representation emphasizes that a strong sparsity of the representation coefficients w is important in representing an input patch, while neglecting the locality constraint which is more essential than sparsity in revealing the true geometry of a nonlinear manifold [43]. In other words, SC may represent one input patch by training data

from irrelevant patches (i.e., patches that are very different from the input patch), which might cause artifacts and incorrect high frequency components in the super-resolved face image.

Motivated by these facts, in our previous work [19], we imposed a locality constraint on the least squares problem and proposed an LcR algorithm to represent an observation LR patch. In the LcR model, for each input LR patch x_t, the reconstruction weights are obtained by the following objective function:

$$J_{\mathrm{LcR}} = \left\| x_t - \sum_{i=1}^{N} w_i x_i \right\|_2^2 + \lambda ||dist \odot w||_2^2, \quad \text{s.t.} \sum_k w_k = 1, \tag{6}$$

where $\odot$ denotes a point wise vector product, λ is a locality regularization parameter which balances the minimization between the reconstruction residual and the locality prior, and *dist* is an N-dimensional locality adaptor that gives different freedom to each training patch x_i according to its similarity to the input LR patch x_t. The similarity is simply measured by the Euclidean distance:

$$dist_i = ||x_t - x_i||_2, \quad 1 \le i \le N. \tag{7}$$

By assigning different freedom to the training samples, LcR is able to achieve sparsity and locality simultaneously.

4. Face SR via Tikhonov regularized neighbor representation (TRNR)

To the best of our knowledge, the LcR based method [19] is currently the best algorithm reported in the literature for face SR, especially under noisy conditions (i.e., the input face image is corrupted by noise). However, LcR [19] represents the input LR patch by all the training patch samples (see Eq. (6)), which will reduce the facial discriminability. For example, the resulting eyes of the reconstructed faces may appear similarly for different person. More faces contributing to the final face SR reconstruction result imply lower discriminability of the reconstructed face. Similar to the idea in traditional neighbor embedding method, in the proposed method we only consider similar faces in the training set as effective samples. Meanwhile, we introduce the Tikhonov regularization term to the neighbor representation objective function, thus leading to (i) a stable solution to the least squares problem and (ii) discriminant HR reconstruction results.

4.1. Objective function of TRNR

In this subsection, we will provide the details of our proposed TRNR model. We introduce the Tikhonov regularization [12], which has been successfully used in compressed sensing image reconstruction [7] and spectral-spatial preprocessing [6], to the ordinary least squares in neighbor embedding to give preference to a particular solution with desirable properties. For each LR patch x_t in the input LR image, the optimal reconstruction weights are obtained by minimizing the Tikhonov regularized reconstruction error:

$$J_{\mathrm{TRNR}}(w) = \left\| x_t - \sum_{k \in C(x_t)} w_k x_k \right\|_2^2 + \lambda ||\Gamma w||_2^2, \quad \text{s.t.} \sum_k w_k = 1, \tag{8}$$

where Γ is the Tikhonov matrix, $C(x_t)$ is the index set of the K-NN of x_t in the LR training patches $\{x_m\}_{m=1}^{M}$, and λ is a global regularization parameter which balances the minimization between the reconstruction error and the Tikhonov regularization term. Considering that different samples should have different freedom, in this paper, we introduce the locality constraint to define the Tikhonov matrix, which has the following form:

$$\Gamma = \begin{bmatrix} ||x_t - x_1||_2 & & 0 \\ & \ddots & \\ 0 & & ||x_t - x_K||_2 \end{bmatrix}. \tag{9}$$

Denote X_K as a matrix with its columns being the K-NN of x_t in the LR space. Thus, Eq. (8) can be written by the following matrix form:

$$J_{\mathrm{TRNR}}(w) = ||x_t - X_K w||_2^2 + \lambda ||\Gamma w||_2^2, \quad \text{s.t.} \sum_k w_k = 1. \tag{10}$$

It is worth noting that we introduce the locality constraint and give different freedom to the training samples for better exploring the data structure. If the Tikhonov matrix is set to an identity matrix, which means all training samples are treated equally, and our proposed method reduces to [17]. From the comparison results (Figs. 3 and 4), we learn that the locality structure of the training samples is important for super-resolving face images. The proposed method can generate better results both in global faces and fine details, especially in face contour and eyes (please refer to the red boxes in Fig. 4).

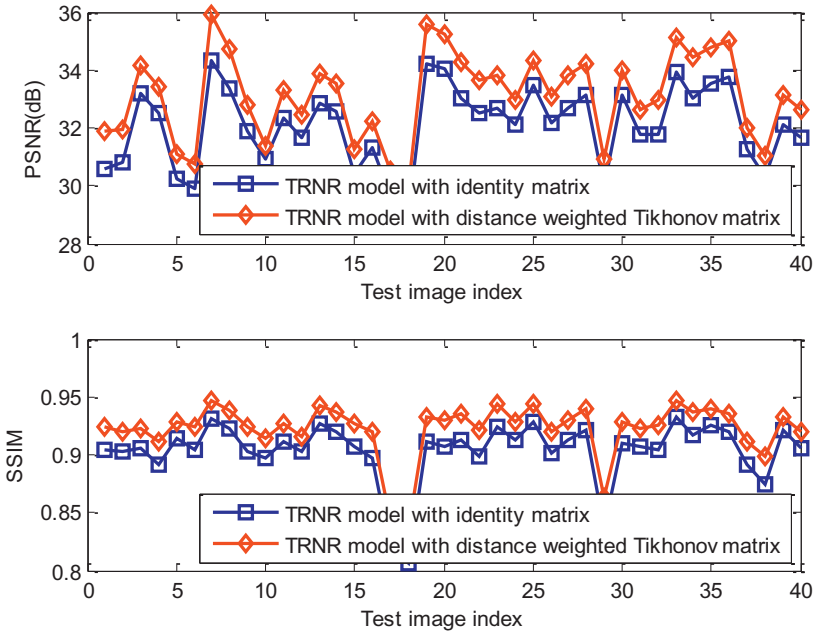


Fig. 3. The PSNR and SSIM [38] comparisons of two different models: TRNR model with identity matrix and TRNR model with distance weighted Tikhonov matrix. The average PSNR and SSIM improvements of the latter over the former are 0.27 dB and 0.0041, respectively. For more details about the experiment setting, please refer to Section 6.

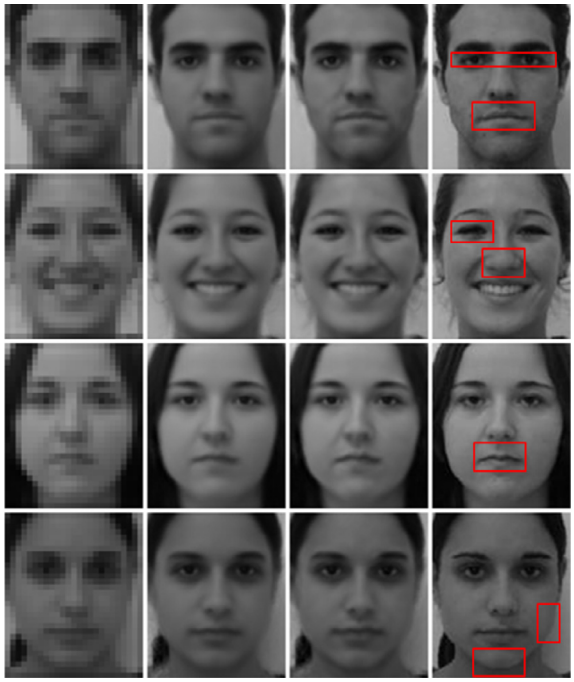


Fig. 4. Visual reconstruction results of two test images of two different models: TRNR model with identity matrix and TRNR model with distance weighted Tikhonov matrix. From left to right: the input LR test faces, super-resolved results of the former, super-resolved results of the latter.

4.2. Optimization of TRNR

In the following, we will present the details of optimizing the TRNR objective function.

Given $\sum_{k=1}^{K} w_k = 1$, we have

$$\begin{aligned} J_{\mathrm{TRNR}}(w) &= \left\| x_t - \sum_{k=1}^{K} x_k w_k \right\|_2^2 + \lambda ||\Gamma w||_2^2 \\ &= \left\| x_t \sum_{k=1}^{K} w_k - \sum_{k=1}^{K} x_k w_k \right\|_2^2 + \lambda ||\Gamma w||_2^2 \\ &= \left\| \sum_{K=1}^{k} (x_t - x_k) w_k \right\|_2^2 + \lambda ||\Gamma w||_2^2 \\ &= \|[x_t - x_1, x_t - x_2, \cdots, x_t - x_K] w\|_2^2 + \lambda ||\Gamma w||_2^2 \\ &= \left\| (x_t \cdot 1^{\mathrm{T}} - X_K) w \right\|_2^2 + \lambda ||\Gamma w||_2^2 \\ &= w^T (x_t \cdot 1^{\mathrm{T}} - X_K)^{\mathrm{T}} (x_t \cdot 1^{\mathrm{T}} - X_K) w + \lambda w^{\mathrm{T}} \Gamma^{\mathrm{T}} \Gamma w \\ &= w^T ((x_t \cdot 1^{\mathrm{T}} - X_K)^{\mathrm{T}} (x_t \cdot 1^{\mathrm{T}} - X_K) + \lambda \Gamma^{\mathrm{T}} \Gamma) w \\ &= w^T Q w, \end{aligned} \tag{11}$$

where $X_K = \{x_k\}_{k \in C_K(x_t)}$, 1 is a $K \times 1$ column vector of ones, "T" denotes matrix transpose, and $Q = (x_t \cdot 1^{\mathrm{T}} - X_K)^T (x_t \cdot 1^{\mathrm{T}} - X_K) + \lambda \Gamma^{\mathrm{T}} \Gamma$.

Under the condition that $\sum_{k=1}^{K} w_k = 1$, Eq. (10) can be written as the following Lagrange form

$$J_{\mathrm{TRNR}}(w, \lambda) = w^T Q w + \lambda (1^{\mathrm{T}} w - 1). \tag{12}$$

Set $\frac{\partial J_{\mathrm{TRNR}}(w,\lambda)}{\partial w} = 0$, we have

$$\frac{\partial J_{\mathrm{TRNR}}(w, \lambda)}{\partial w} = 2Qw + \lambda 1 = 0. \tag{13}$$

$$w = -\frac{\lambda}{2} Q^{-1} 1. \tag{14}$$

Set $\frac{\partial J_{\mathrm{TRNR}}(w,\lambda)}{\partial \lambda} = 0$, we have

$$\frac{\partial J_{\mathrm{TRNR}}(w, \lambda)}{\partial \lambda} = 1^{\mathrm{T}} w - 1 = 0. \tag{15}$$

From Eqs. (14) and (15), we obtain

$$1^{\mathrm{T}} w = 1^{\mathrm{T}} \left(-\frac{\lambda}{2} Q^{-1} 1 \right) = -\frac{\lambda}{2} 1^{\mathrm{T}} Q^{-1} 1 = 1. \tag{16}$$

Thus,

$$-\frac{\lambda}{2} = (1^{\mathrm{T}} Q^{-1} 1)^{-1}. \tag{17}$$

From Eqs. (14) and (17), we can get the final solution to Eq. (12):

$$w = \frac{Q^{-1} 1}{1^{\mathrm{T}} Q^{-1} 1}. \tag{18}$$

To avoid the inversion of Q, a more efficient way is to solve the linear system of equations $Qw = 1$, and then normalize the weights so that $\sum_{k=1}^{K} w_k = 1$. Thus,

$$w = Q \backslash 1, \tag{19}$$

where "\" denotes the left matrix division. For example, $A \backslash B$ is the matrix division of A into B, which is roughly the same as $A^{-1}B$.

4.3. Face SR via TRNR

As a local patch based face SR method, the input face image is first decomposed into small patches according to their positions. The patches are processed in the raster-scan order, from left to right and top to bottom. With the manifold assumption, which states that the LR and HR patch pair share similar local geometry structure (i.e., the reconstruction weights

in image SR), the target HR image patch can be generated by applying the same reconstruction weights to the corresponding neighbor HR patches in the HR space:

$$y_t = \sum_{k \in C(x_t)} w_k y_k. \tag{20}$$

Finally, following [5,19], the compatibility between adjacent patches is enforced by averaging pixel values in the overlapping regions. The steps of TRNR are summarized in Algorithm 1.

Algorithm 1 Face SR via TRNR.

Input: Training sets of LR and HR face images, $I^L = \{I_1^L, I_2^L, \cdots, I_N^L\}$ and $I^H = \{I_1^H, I_2^H, \cdots, I_N^H\}$, an input LR face image I_t^L, *patch_size*, *overlap*, K, and λ.
Output: HR face image I_t^H .
Compute U and V:
$U = \text{ceil}\left\{\frac{imrow - overlap}{patch\_size - overlap}\right\}$
$V = \text{ceil}\left\{\frac{imcol - overlap}{patch\_size - overlap}\right\}$
Divide each of the LR and HR training images and the input LR image into N small patches to form patch sets: $\{x_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$, $\{y_i(p,q)|1 \le p \le U, 1 \le q \le V\}_{i=1}^N$ and $\{y_t(p,q)|1 \le p \le U, 1 \le q \le V\}$.
for p = 1 to U **do**
  for q = 1 to V **do**
    Compute the distance between $x_t(p,q)$ and $x_i(p,q)$, $i = 1, 2, \ldots, N$.
    Obtain the index set $C(x_t)$.
    Obtain the K-NN set X_K.
    Calculate the matrix Q.
    Compute the representation coefficients w by solving $Qw = 1$, s.t. $\sum_k w_k = 1$.
    Reconstruct the target HR image patch $y_t = \sum_{k \in C(x_t)} w_k y_k$.
  end for
end for
Integrate all the obtained HR patches according to their positions. The final HR image I_t^H can be generated by averaging pixel values in the overlapping regions.

5. Connections to existing position-patch based methods

Note that our method is similar to the neighbor embedding method [5] and the LcR method [19], which learn the LR and HR patches relationship by neighbor embedding or exploring the locality prior of the training space. However, our proposed method differs from these approaches in the following:

- Our proposed method inherits the merits of neighbor embedding strategy by exploring the neighbor relationship (reconstruction weights) in the LR space and embedding this relationship into the HR space. However, there are essential differences between the traditional neighbor embedding and our proposed TRNR method. Traditional neighbor embedding based image SR method uses a fixed K neighbors and assigns the same weights to these neighbors for reconstruction, which often results in blurring effects. We attribute this disadvantage to the over- or under-fitting of the linear combination problem of the selected K neighbors.
- The LcR method incorporates the local geometrical constraint of manifold into the patch reconstruction objective function to achieve sparsity and locality simultaneously. Since it represents the input LR patch image collaboratively over all the training patches, it may also "*synthesize*" the input noisy. Inspired by searching and weighting strategies in nonlocal means denoising method, we propose to select only similar training patches (discard the dissimilar training patches) and assign different weights to them.

6. Experiments and result analysis

In this section, we describe the details of the experiments performed to evaluate the generality, effectiveness, and robustness of the proposed method for face SR. We also compare our proposed method with some recent state-of-the-art methods. The experiments are carried out on the public FEI face database[2] [33]. Two quality measures, PSNR and SSIM index are adopted to evaluate the objective quality of the reconstructed results. We evaluate our proposed TRNR method

[2] http://fei.edu.br/~cet/facedatabase.html

Table 1
The optimal K and λ of NE [5], LSR [29], LcR [19] and our proposed TRNR methods. Note that the underlined values are fixed.

Noise levels	NE		LSR		LcR		TRNR	
	K	λ	K	λ	K	λ	K	λ
$\sigma = 0$	75	0	360	0	360	0.01	150	0.01
$\sigma = 10$	5	0	360	0	360	10	50	1
$\sigma = 20$	1	0	360	0	360	10	20	1
$\sigma = 25$	1	0	360	0	360	10	20	1
$\sigma = 30$	1	0	360	0	360	10	20	10

and some representative algorithms, such as Wang et al. [37]'s eigentransformation method, Chang et al.'s NE method [5], Ma et al.'s LSR method [29], Yang et al.'s SC [40] method, and our previously proposed LcR method [19]. Note that the results of the Bicubic interpolation method are given as baselines for comparison. Matlab code available upon e-mail request (junjun0595@163.com).

6.1. Database description

The FEI face database consists of 400 facial images. All the images are cropped to 120 $\times$ 100 pixels to form the HR training faces. Human faces in the database are mainly from 19 to 40 years old with distinct appearances, e.g., hairstyles and adornments. The LR images are formed by smoothing (by a 4 $\times$ 4 mean filter) and down-sampling (by a factor of 4 resulting the size of LR face images to be 30 $\times$ 25 pixels) the corresponding HR images, and then adding additive white Gaussian noise (AWGN) with different noise levels (denoted by σ). In our experiments, we use 360 images to train the proposed algorithm, leaving the rest 40 images for testing. Therefore, all the test images are absent completely in the training set. Note that the face can be aligned by some recently proposed automatic alignment methods [27] and feature points matching methods [28].

6.2. Parameter settings

All parameter settings reported in this paper are experimentally determined. In order to achieve the optimal performance, an SR method may take different parameters at different levels of noise. Particularly, for Wang et al. [37]'s global face based eigentransformation method, we let the variance accumulation contribution rate of PCA be 0.999 ($\sigma = 0$), 0.99 ($\sigma = 10$), 0.97 ($\sigma = 20$), 0.95 ($\sigma = 25$), and 0.95 ($\sigma = 30$), respectively. For the patch based methods, e.g., NE [5], LSR [29], SC [40], LcR [19] and our proposed TRNR, the patch size and overlap between neighbor patches are particularly important for obtaining reliable results. If the patch size is too small, the co-occurrence geometric structure information in LR and HR spaces is too weak to make the prediction meaningful. On the contrary, if the patch size is too large (especially when the test image is very different from the training samples), not only a large training set is needed to find similar patches for the test image, but also the reconstructed HR face image may be smoothed and lose some visual details. According to [19], in this paper we set patch size and overlap to the values when the performance of the position-patch based face SR methods reach the maximum. In all our experiments, the patch size is 16 $\times$ 16 pixels for HR patches and the overlap between neighbor patches is 12 pixels, while the corresponding LR patch size is set to 4 $\times$ 4 pixels with an overlap of 3 pixels. In addition to patch size and overlap between neighbor patches, the regularization parameter and the neighbor number are important for the position-patch based representation models. As for the error tolerance in the SC method [40], we set it to 0.005 ($\sigma = 0$), 0.1 ($\sigma = 10$), 0.2 ($\sigma = 20$), 0.3 ($\sigma = 25$), and 0.3 ($\sigma = 30$), respectively. As discussed in *Remark 3* of Section 4.3, NE [5], LSR [29] and LcR [19] can be seen as different cases of our proposed TRNR method. For comparison, we report the optimal parameter settings (i.e., the neighbor number K and the regularization parameter λ) in Table 1. Note that λ is fixed to 0 in NE [5], K and λ are fixed to 360 and 0 in LSR [29], and K is fixed to 360 in [19]. We use underlined values to distinguish these fixed values from the model parameters of different methods. From Table 1, with the increase of noise level, the neighbor number K decreases while the regularization parameter λ increases for our proposed TRNR method. This is mainly because that in the SR reconstruction of a noisy LR image, we aim to recover the latent HR image patches with similar *clean* ones, not *synthesize* and *explain* the noisy patch by all the training samples. It is also worth noting that the number of selected patches is small (sparsity) and the selected patches are similar to the input one (locality), which are the two characteristics of our proposed TRNR method discussed in *Remark 2* of Section 4.3. For more detailed analysis about the parameter settings with different noise levels, please refer to the following subsection.

6.3. Effectiveness of the Tikhonov regularization

To verify the effectiveness of Tikhonov regularization in our proposed TRNR algorithm, in the following, we evaluate the performance of TRNR according to different parameter settings, i.e., neighbor number K and Tikhonov regularization parameter λ, under different noise levels (e.g., $\sigma = 0, 10, 20, 25, 30$).

Fig. 5 plots the PSNR and SSIM results according to K and λ when σ is set to 0, 10, 20, 25 and 30, respectively. We find in this experiment that (i) in addition to K, the regularization parameter λ also has a great influence on the face SR

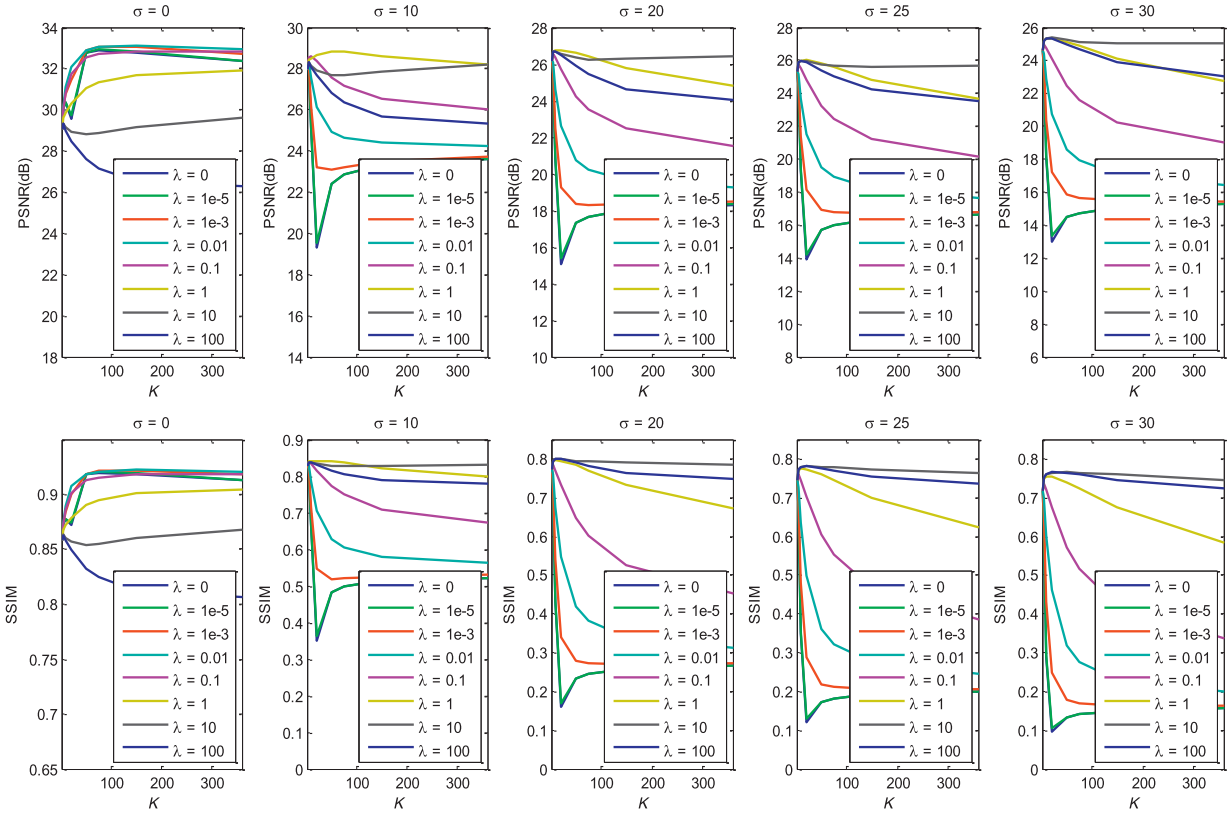
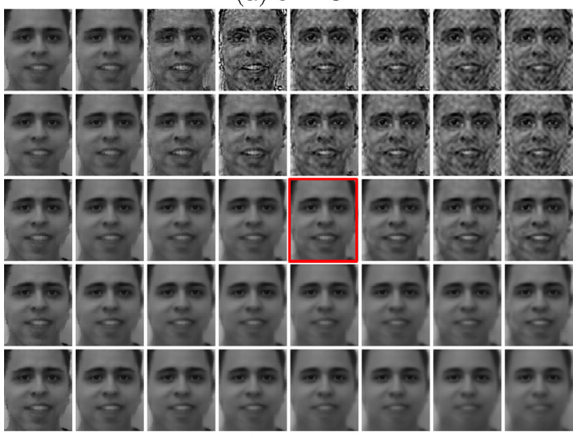


Fig. 5. The PSNR and SSIM results of the proposed TRNR according to K and λ when σ is set to 0, 10, 20, 25, and 30, respectively.

performance. This implies the effectiveness of Tikhonov regularization in neighbor embedding; (ii) when the value of λ is set to zero, our proposed TRNR method reduces to the traditional NE method [5]. We can see from Fig. 5 that when the input is clean ($\sigma = 0$), the performances of NE and TRNR are substantially equivalent. However, when the noise level becomes large, by setting a large value of λ, the performance gain of TRNR over NE is remarkable; (iii) when the value of K is set to 360, our proposed TRNR method reduces to the LcR method [19]. We can see from Fig. 5 that when the input is disturbed by high level noise, e.g., $\sigma = 30$, larger K will yield worse performance. This verifies the significance of selecting several similar patches for robust representation of the noisy patch; (iv) when the input image has strong noise corruption, it needs a relatively large regularization parameter λ to balance the data term and the Tikhonov regularization term, which further indicates the effectiveness of Tikhonov regularization, especially when the noise level is very high.

In addition to the PSNR and SSIM performances shown in Fig. 5, we also present some visual reconstruction results of our proposed TRNR in Figs. 6 and 7. For each subfigure in Fig. 6, from left to right, they are the reconstruction results for $K = 1$, 5, 10, 20, 50, 75, 150, and 360, respectively; from top to bottom, they are the reconstruction results for $\lambda = 0$, 0.01, 1, 10, and 100, respectively. We highlight the optimal results of our proposed TRNR by red boxes. Fig. 7 demonstrates the corresponding PSNRs and SSIMs of the reconstructed HR faces. Due to limited space, we only show several results under typical situations. We learn from these results that (i) the neighbor number K and Tikhonov regularization parameter λ play significant roles in super-resolving the noise input LR images. By choosing appropriate parameters, our proposed TRNR method can well maintain the facial structure and remove most of the added noise. Please see the optimal reconstruction results marked by red boxes in Fig. 6. (ii) when K is large and λ is small (e.g., $K > 50$ and $\lambda < 0.01$), TRNR will make use of the training samples to "synthesize" the input LR image as well as the input noise, thus recovering a noisy HR image due to over-fitting (see the subplots in the top right corners of Fig. 6(b) and (c)); when K and λ both are set to large values (e.g., $K > 75$ and $\lambda > 10$), TRNR will treat the selected training samples equally and produce mean face like results (see the bottom right corners of each sub-figure in Fig. 6). (iii) an efficient way of setting the parameters is to let K to be relatively small and let λ to be relatively large, especially when the input LR is corrupted by noise (e.g., when $\sigma = 10$, K should be less than 75 and λ should be larger than 0.01; when $\sigma = 30$, K should be less than 50 and λ should be larger than 1). This point can be concluded from Table 1 as well; (iv) the large performance drop and instability occur when $K = 10$ (see the third column of each sub-figure in Fig. 6). This is because in our experiments, the LR image patch is 4×4 (16-dimensional feature vector). If K is set to around 16, the least squares system will be very instable and *over fitted* on the observation data. At the same time, we also observe that this shortcoming can be overcame by introducing the Tikhonov regularization term and giving a

(a) $\sigma = 0$



(b) $\sigma = 10$

(c) $\sigma = 30$

Fig. 6. Visual reconstruction results of TRNR with different levels of noise observation according to various neighbor numbers K and Tikhonov regularization parameters λ.

relatively large weight to this term to balance the instability of the least squares system (see the objective function (8) of TRNR). For visual results, see the third to the fifth rows of the third column in each sub-figure of Fig. 6.

6.4. Comparisons with state-of-the-art approaches

Table 2 tabulates the comparisons of PSNR and SSIM of different SR methods with different noise levels. To make the comparison more clear, we also show the PSNR and SSIM improvements of TRNR over the five comparison methods in

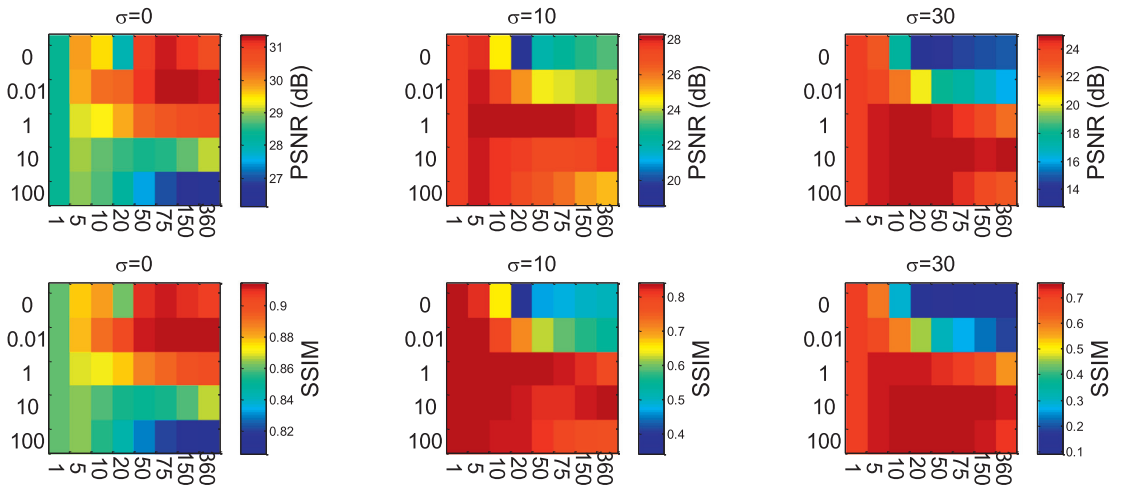


Fig. 7. The objective results in terms of PSNR and SSIM corresponding to the visual reconstruction results in Fig. 6 by various combinations of the parameters K (the horizontal axis) and λ (the vertical axis).

Table 2
PSNR (dB)and SSIM comparisons of seven different methods: Bicubic, Wang et al.'s method [37], NE [5], LSR [29], SC [40], LcR [19], and TRNR. Red indicates the best and blue indicates the second best performance.

Noise levels	Bicubic	Wang	NE	LSR	SC	LcR	TRNR
$\sigma = 0$	27.50	27.57	32.89	32.36	32.77	32.97	33.12
	0.8426	0.7453	0.9197	0.9132	0.9184	0.9208	0.9222
$\sigma = 10$	25.41	26.68	28.13	23.60	27.87	28.18	28.82
	0.6856	0.7495	0.8178	0.5206	0.7782	0.8323	0.8410
$\sigma = 20$	22.16	25.32	26.14	18.32	24.09	26.41	26.77
	0.4742	0.7303	0.7728	0.2657	0.6073	0.7839	0.7947
$\sigma = 25$	20.73	24.97	25.30	16.62	24.11	25.68	26.00
	0.3928	0.7340	0.7438	0.2002	0.6375	0.7635	0.7733
$\sigma = 30$	19.44	24.60	24.57	15.26	21.38	25.06	25.36
	0.3269	0.7252	0.7180	0.1557	0.4683	0.7458	0.7646

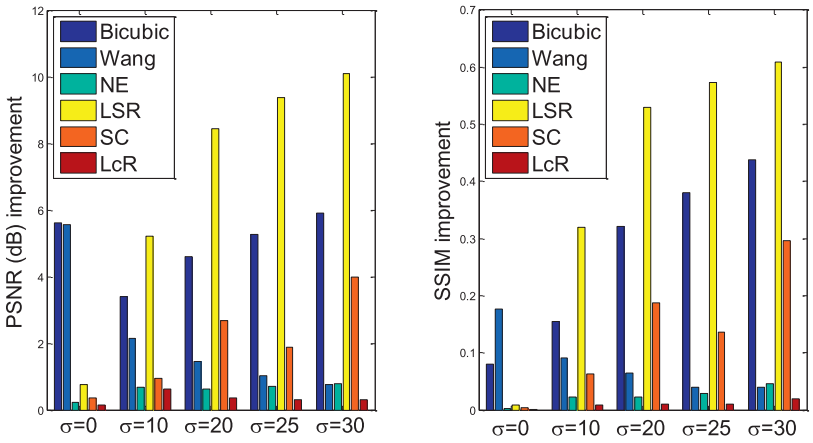


Fig. 8. The PSNR(dB) and SSIM improvements under different noise levels over the five comparison methods.

Fig. 8. We find that the proposed TRNR method achieves the best performance regardless of noise levels. Moreover, with the increase of the noise level, the gain of TRNR over these position-patch based approaches, e.g., NE [5], LSR [29] and SC [40], becomes increasingly apparent. The improvement of TRNR over the Bicubic interpolation firstly decreases, and then increases. This is mainly because the Bicubic interpolation can smooth some small noise (e.g., $\sigma = 10$). It is also found that the improvement of TRNR over [37] is becoming increasingly insignificant due to the noise robust PCA model.

Fig. 9. Super-resolved faces of two LR test samples with different noise levels (from top to bottom, the noise levels are $\sigma = 0$, $\sigma = 10$, $\sigma = 20$, $\sigma = 25$, and $\sigma = 30$, respectively) of different methods (from left to right, the input LR images, Bicubic interpolation, Wang et al. [37], NE [5], LSR [29], SC [40], LcR [19], our proposed TRNR, and the ground truth HR face images.).

Due to space limitation, we only exhibit the results of two test images in Fig. 9, the performances of the position-patch based methods are almost the same. These methods can well recover the detailed face features and maintain the face contour information. As a global face based method, Wang et al.'s method [37] has severe "ghosting" effects and causes serious distortions, especially around face contours (see the first two rows of the third column in Fig. 9). This is mainly because that [37] is based on a PCA statistical mode, which can not reveal the distribution of data when the training set size is not large enough (i.e., 360) as compared to the data dimension (i.e., 750, which is the size of the input LR image (30×25)).

As the noise becomes large, LSR [29] can hardly reconstruct reasonable faces (see the fifth column in Fig. 9). The is due to the unstable solution of solving the least squares problem in [29]. When the noise level is high, e.g., $\sigma > 10$, the SC method [40] will cause severe distortion (see the last six rows of the sixth column in Fig. 9). This is consistent with the sparse recovery theory which states that sparse coding is robust against small magnitude noise in the observation [11]. As shown in the third column, Wang et al.'s method [37] can well maintain the primitive facial feature information and remove most of the noise, but the reconstructed faces are dirty and different from the ground truth especially when the noise increases. It is observed that NE method [5] can remove the noise, but the results have obvious blocking artifacts.

Fig. 10. Images captured by a surveillance camera at different distances and lighting intensities. The third row can be seen as the "ground truth" that captured in a normal light and near the camera condition.

This is because it chooses a fixed number of K, which will lead to an over- or under-fitting solution. The super-resolved HR faces by LcR [19] are smooth and lack of details when compared with our proposed method (please see the noses and face contours of the seventh and eighth columns in Fig. 9.). We attribute this to the fact that LcR [19] unitizes all the training samples to represent the observation, thus lacking discriminant facial features. However, in the proposed method, we select similar samples and exclude the impact of "*unrelated*" samples, thus enhancing the discriminant of the reconstructed HR faces and recovering more face features. Simulation experiments demonstrate that our approach is able to generate HR face images with visually satisfactory global face appearance and local detailed features.

6.5. SR with surveillance images

The image degradation process including the additive noise is under simulation conditions. The LR training samples are formed by simply smoothing, down-sampling and adding noise to the HR training samples, and the co-occurrence prior is applied to regularize the ill-posed image SR reconstruction problem. Such a prior learned from the simulated training samples cannot represent the true spatial relationship between LR and HR pairs as they do not correspond to the images captured by a real camera [10]. In real world conditions, it is too difficult for us to simulate the image degradation process.

In order to further test and verify the effectiveness of our method under the real surveillance imaging condition, we conduct experiments on some surveillance images as shown in Fig. 10. The surveillance images are CIF-size (352 $\times$ 288 pixels).

In our experiments, we firstly calculate the mean face of the FEI face database, and then manually adjust the size of the extracted LR face from the captured picture and align the extracted face to the mean face of the FEI face database. Then, the RGB face is adjusted by the histogram and converted to gray level. We manually extract and align the input faces in the captured images according to the two points of eye centers, and then crop them to 30 $\times$ 25 pixels (the first row), 60 $\times$ 50 pixels (the second row), and 120 $\times$ 100 pixels (the third row) as shown in Fig. 10. Then, we convert the color images to gray-scale and adjust the intensity levels as shown in the first column of Fig. 11. We use the same 360 training samples from the FEI face database as described in Section 6.1 to train the TRNR model. The second to the eighth columns of Fig. 11 are the super-resolved HR face images, whereas the last column is the "ground truth" for visual comparisons. From the above experiments, we can see that the Bicubic interpolation and LSR [29] methods are almost impossible to remove the noise, the super-resolved results of [37] are similar to a mean face or have "ghost effect", SC [40] obtains dirty faces, and locality regularized methods (LcR [19] and the proposed TRNR) can remove most of the noise and maintain the main contour information.

When compared with the results on the FEI face database (Fig. 9), the real surveillance results (Fig. 11) appear to be artificial and are relatively poor. This is due to the inconsistent statistics between the training samples (the FEI face database) and the observations (surveillance face images), indicating a disadvantage of the learning-based SR techniques which require a certain similarity between the training and testing images [26].

7. Discussion

The effect of the Tikhonov regularization term in the objective function (10) is twofold. First, if the coherence between samples is large (i.e., training samples are sufficiently similar to each other) or the training set is overcomplete (i.e., the size of a training set is much larger than the dimension of a sample), the inverse problem of SR reconstruction will either have poor conditioning or be near-singular. Therefore, the patch representation will have high variance and convey little to no meaning. By adding the Tikhonov regularization term to the patch representation objective function, we can improve

Fig. 11. Super-resolved results of surveillance images of different methods (from left to right, the input LR images, Bicubic interpolation, Wang et al. [37], NE [5], LSR [29], SC [40], LcR [19], our proposed TRNR, and the "ground truth" HR face images.

Table 3
The GI of five different position-patch based patch representation methods: NE [5], LSR [29], SC [40], LcR [19] and TRNR.

Methods	NE	LSR	SC	LcR	TRNR
GI	0.2250	0.1670	0.0.4481	0.2280	0.2831

the conditioning of the ill-posed SR problem. Second, by utilizing the Tikhonov regularization term with a distance weighting strategy, the TRNR model is able to assign larger reconstruction coefficients to the samples that are most similar to the testing sample than the samples that are most dissimilar. It can effectively enhance the discriminative power of the reconstructed HR face image.

Recent studies indicate that sparsity and locality priors are critical in computer vision and machine learning tasks [31,43]. To achieve satisfactory SR reconstruction performance, the solution of the position-patch representation based methods should satisfy the properties of sparsity and locality. Here we compare the sparsity and locality of NE [5], LSR [29], SC [40], LcR [19] and our proposed TRNR using two measurements, e.g., gini index (GI) [47] and K-mean distance (K-MD) [19]. GI is normalized, and assumes values between 0 and 1 for any vector. Moreover, it is 0 for the least sparse signal with all the coefficients having an equal amount of energy; and 1 for the sparsest one which has all the energy concentrated in just one coefficient. K-MD is the mean Euclidean distance between the input LR patch and K largest weights training patches. For more details about these two measurements, please refer to the reference therein. We conduct experiments on the FEI face Database [33] (more details about the database are given in Section 6.1). From Table 3, SC reaches the most sparse solution among all the comparison methods. With the locality constraint, NE [5], LcR [19] and our proposed TRNR all obtain sparse results to some extent compared with LSR [29]. Fig. 12 plots the histograms of the K-MD for different methods. TRNR obtains the most locality results regardless of the values of K, while the results of LSR [29] and SC [40] do not have this

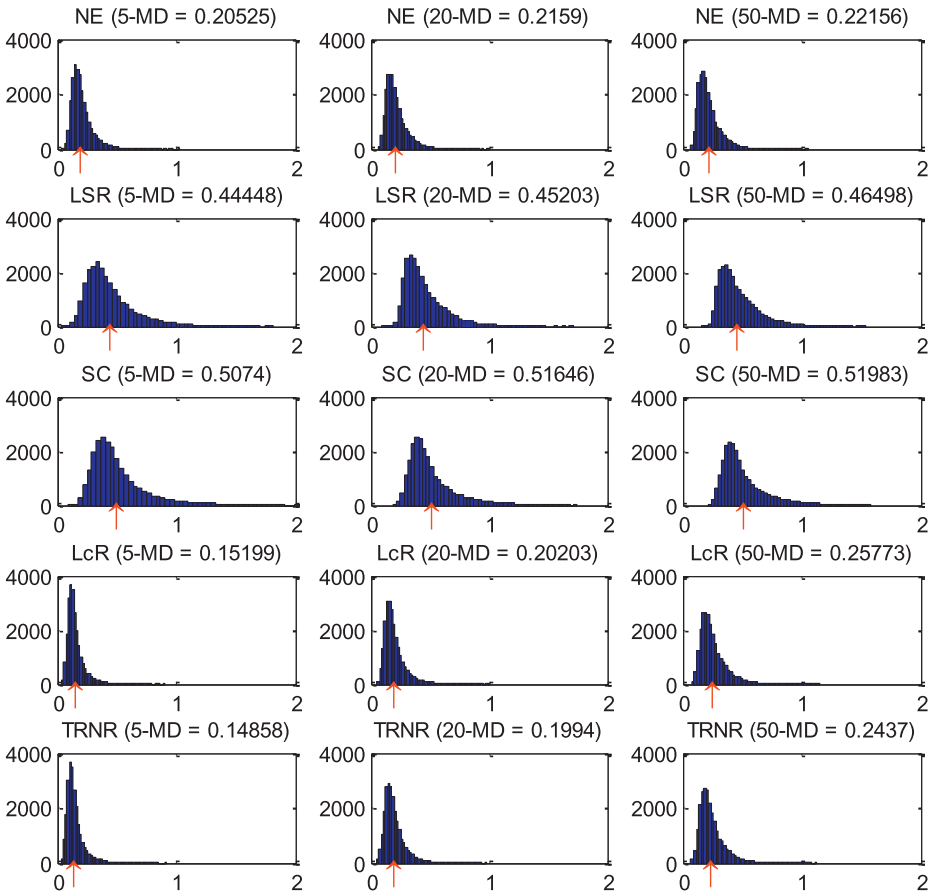


Fig. 12. Histograms of the K-MD for different methods when K is set to some typical values, e.g., $K = 5, 20, 50$. Note that the red arrows indicate the average values, and we also show the mean K-MD in the sub-caption of different methods with different K.

property. In addition, with the increase of K, the K-MD indices of NE [5], LcR [19] and our proposed TRNR also increase, while the K-MD indices of LSR [29] and SC [40] change randomly. This implies that LSR [29] and SC [40] may give large weight coefficients to the training samples that are dissimilar from the observation. Therefore, LSR [29] and SC [40] fail to explore the locality prior in the training data.

From the objection function of NE [5], LSR [29], SC [40], LcR [19] and our proposed TRNR, we learn that when $\lambda = 0$, the TRNR model reduces to the traditional NE approach [5]; when K equal to the training set size, the TRNR model reduces to our previously proposed LcR approach [19]; when $\lambda = 0$ and K equal to the training set size, the TRNR model is equivalent to Ma et al.'s LSR based approach [29]. To achieve the optimal performance of our proposed TRNR algorithm, we change the value of K (from 1 to the training set size) and then adjust the value of λ. We find that this process resembles the matching pursuit algorithm [35] in sparse recovery, where the algorithm first selects the nearest atom and then subtracts the contribution of that atom, and repeats the process until the observation is satisfactorily reconstructed. Therefore, our proposed TRNR model can be seen as the generalization of those four methods mentioned above.

We currently use the frontal faces as the training set, only when the input LR is frontal face or near frontal face, we can hope to obtain a satisfactory result. This is demonstrated by the real world experiments. When the input LR face image confronts with low-light or pose variety, the super-resolved result is often worse than that in FEI face database. If we can pre-construct a co-occurrence model between the LR training set with arbitrary pose and the corresponding HR training set with frontal pose, we can expect to infer a good result. But in practice, we cannot get face sample with arbitrary pose. When confronted with face image at an angle, the general practice is to assign the input face to the nearest pose in the training set, and then perform SR reconstruction as the frontal case. In the actual monitoring environment, we often can get more than one face image, such as a sequence. Therefore, in addition to the prior information learned from the co-occurrence model, the temporal redundancy information can be also used to guide the reconstruction process. Combining with the advantages of learning-based methods (explore the prior information from the given training set) and reconstruction-based methods (explore the prior information from the given sequence or features of different views by multiview learning methods [14,42]) to recover a high-quality and HR face images will be our further work, and this also make it possible to solve the occlusion

problem. As for the illumination problem, we can use some pre-processing techniques and extract some illumination robust features to alleviate the negative effect.

8. Conclusion and future work

In this paper, we have proposed a novel position-patch based face SR method via Tikhonov regularized neighbor representation (TRNR). By incorporating the Tikhonov regularization term to the neighbor embedding, TRNR can select the most relevant patch samples (especially when the face image is corrupted by noise) to reconstruct the HR version of the observed LR image patch, thus generating discriminant HR face image with detailed features. Experimental results demonstrate the generality, effectiveness, and robustness of the proposed algorithm.

In this paper, we have focused on designing a frontal face SR method. However, when the LR observation face is in the wild, i.e., arbitrary pose, various skin colors, and extreme ambient illumination, how to recover the frontal HR face image is our future work. By decomposing a complete face image into small face patches, we can expect to enhance the representation ability of the training set. However, this will lead to a sharp increase in algorithm complexity. In order to promote practical applications of our algorithm, accelerating the algorithm via parallel computation or using graphics processing unit (GPU) will be another direction of our future work.

Acknowledgment

The research was supported by the National Natural Science Foundation of China under Grant 61501413, the Fundamental Research Funds for the Central Universities at China University of Geosciences (Wuhan) under Grant CUGL160412, and the Scientific Research Project of Hubei Provincial Department of Education under Grant Q20152907.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ins.2016.05.032

References

[1] L. An, B. Bhanu, Face image super-resolution using 2D CCA, Signal Process. 103 (2014) 184–194.
[2] S. Baker, T. Kanade, Hallucinating faces, in: Proc. IEEE Conf. on Automatic Face and Gesture Recognition, 2000, pp. 83–88.
[3] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2, IEEE, 2005, pp. 60–65.
[4] A. Chakrabarti, A. Rajagopalan, R. Chellappa, Super-resolution of face images using kernel PCA-based prior, IEEE Trans. Multimedia 9 (4) (2007) 888–892.
[5] H. Chang, D. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1, 2004, pp. 275–282.
[6] C. Chen, W. Li, E.W. Tramel, M. Cui, S. Prasad, J.E. Fowler, Spectral-spatial preprocessing using multihypothesis prediction for noise-robust hyperspectral image classification, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7 (4) (2014) 1047–1059.
[7] C. Chen, E.W. Tramel, J.E. Fowler, Compressed-sensing recovery of images and video using multihypothesis predictions, in: Proceedings of the 45*th* Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, 2011, pp. 1193–1198.
[8] C. Dong, X. Zhu, Y. Deng, C.C. Loy, Y. Qiao, Boosting optical character recognition: a super-resolution approach, arXiv preprint arXiv:1506.02211(2015).
[9] W. Freeman, E. Pasztor, O. Carmichael, Learning low-level vision, Int. J. Comput. Vis. 40 (2000) 25–47.
[10] P. Gajjar, M. Joshi, New learning based super-resolution: use of dwt and igmrf prior, IEEE Trans. Image Process. 19 (5) (2010) 1201–1213.
[11] J.F. Gemmeke, T. Virtanen, A. Hurmalainen, Exemplar-based sparse representations for noise robust automatic speech recognition, IEEE Trans. Audio Speech Lang. Process. 19 (7) (2011) 2067–2080.
[12] G.H. Golub, P.C. Hansen, D.P. O'Leary, Tikhonov regularization and total least squares, SIAM J. Matrix Anal. Appl. 21 (1) (1999) 185–194.
[13] W. Gong, L. Hu, J. Li, W. Li, Combining sparse representation and local rank constraint for single image super resolution, Inf. Sci. 325 (2015) 1–19.
[14] C. Hong, J. Yu, J. You, X. Chen, D. Tao, Multi-view ensemble manifold regularization for 3D object recognition, Inf. Sci. 320 (2015) 395–405.
[15] Y. Hu, K.-M. Lam, G. Qiu, T. Shen, From local pixel structure to global image super-resolution: a new face hallucination framework, IEEE Trans. Image Process. 20 (2) (2011) 433–445.
[16] H. Huang, H. He, X. Fan, J. Zhang, Super-resolution of human face image using canonical correlation analysis, Pattern Recogn. 43 (7) (2010) 2532–2543.
[17] K. Huang, R. Hu, J. Jiang, Z. Han, F. Wang, Cdmma: coupled discriminant multi-manifold analysis for matching low-resolution face images, in: Proc. IEEE Conf. on MultiMedia Modeling, 2016, pp. 409–420.
[18] J. Jiang, R. Hu, Z. Wang, Z. Han, Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning, IEEE Trans. Image Process. 23 (10) (2014) 4220–4231.
[19] J. Jiang, R. Hu, Z. Wang, Z. Han, Noise robust face hallucination via locality-constrained representation, IEEE Trans. Multimedia 16 (5) (2014) 1268–1281, doi:10.1109/TMM.2014.2311320.
[20] J. Jiang, R. Hu, Z. Wang, Z. Han, J. Ma, Facial image hallucination through coupled-layer neighbor embedding, IEEE Trans. Circuits Syst. Video Technol. PP (99) (2015) 1–1, doi:10.1109/TCSVT.2015.2433538.
[21] J. Jiang, X. Ma, Z. Cai, R. Hu, Sparse support regression for image super-resolution, IEEE Photon. J. 7 (5) (2015) 1–11.
[22] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, in: Advances in Neural Information Processing Systems, 2007, pp. 801–808.
[23] J. Li, W. Gong, W. Li, Dual-sparsity regularized sparse representation for single image super-resolution, Inf. Sci. 298 (2015) 257–273.
[24] K. Li, J. Yang, J. Jiang, Non-rigid structure from motion via sparse representation, IEEE Trans. Cybern. 45 (8) (2015) 1401–1413.
[25] Y. Liang, X. Xie, J.-H. Lai, Face hallucination based on morphological component analysis, Signal Process. 93 (2) (2013) 445–458.
[26] C. Liu, H.-Y. Shum, C.-S. Zhang, A two-step approach to hallucinating faces: global parametric model and local nonparametric model, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1, 2001, pp. 192–198.
[27] J. Ma, W. Qiu, J. Zhao, Y. Ma, A.L. Yuille, Z. Tu, Robust l_2e estimation of transformation for non-rigid registration, IEEE Trans. Signal Process. 63 (5) (2015) 1115–1129.
[28] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, J. Tian, Robust feature matching for remote sensing image registration via locally linear transforming, IEEE Trans. Geosci. Remote Sens. 53 (12) (2015) 6469–6481.

[29] X. Ma, J. Zhang, C. Qi, Hallucinating face by position-patch, Pattern Recogn. 43 (6) (2010) 2224–2236.
[30] J.-S. Park, S.-W. Lee, An example-based face hallucination method for single-frame, low-resolution facial images, IEEE Trans. Image Process. 17 (10) (2008) 1806–1816.
[31] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[32] J. Shi, X. Liu, C. Qi, Global consistency, local sparsity and pixel correlation: a unified framework for face hallucination, Pattern Recognit. 47 (11) (2014) 3520–3534.
[33] C.E. Thomaz, G.A. Giraldi, A new ranking method for principal components analysis and its application to face image analysis, Image Vis. Comput. 28 (6) (2010) 902–913, doi:10.1016/j.imavis.2009.11.005.
[34] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc., Ser. B 58 (1994) 267–288.
[35] J. Tropp, A.C. Gilbert, et al., Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.
[36] N. Wang, D. Tao, X. Gao, X. Li, J. Li, A comprehensive survey to face hallucination, Int. J. Comput. Vis. 106 (1) (2014) 9–30.
[37] X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev. 35 (3) (2005) 425–434.
[38] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[39] Z. Wang, R. Hu, S. Wang, J. Jiang, Face hallucination via weighted adaptive sparse regularization, IEEE Trans. Circuits Syst. Video Technol. 24 (5) (2014) 802–813.
[40] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[41] J. Yu, R. Hong, M. Wang, J. You, Image clustering based on sparse patch alignment framework, Pattern Recognit. 47 (11) (2014) 3512–3519.
[42] J. Yu, Y. Rui, Y.Y. Tang, D. Tao, High-order distance-based multiview stochastic learning in image classification, IEEE Trans. Cybern. 44 (12) (2014) 2431–2442, doi:10.1109/TCYB.2014.2307862.
[43] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: Advances in Neural Information Processing Systems, 2009, pp. 2223–2231.
[44] K. Zeng, J. Yu, R. Wang, C. Li, D. Tao, Coupled deep autoencoder for single image super-resolution, IEEE Trans. Cybern. In Press (2015).
[45] W. Zhang, W.-K. Cham, Hallucinating face in the DCT domain, IEEE Trans. Image Process. 20 (10) (2011) 2769–2779.
[46] Y. Zhu, K. Li, J. Jiang, Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow, Signal Processing: Image Communication 29 (8) (2014) 875–886.
[47] D. Zonoobi, A. Kassim, Y.V. Venkatesh, et al., Gini index as sparsity measure for signal reconstruction from compressive samples, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 927–932.
[48] K. Li, Y. Zhu, J. Yang, J. Jiang, Video super-resolution using an adaptive superpixel-guided auto-regressive model, Pattern Recognit. 51 (2016) 59–71.