

Parameter estimation in sparse representation based face hallucination



Zhongyuan Wang*, Ruimin Hu, Zhenfeng Shao, Zhiqiang Hou

NERCMS, School of Computer, Wuhan University, Wuhan, 430072, China

ARTICLE INFO

Article history:

Available online 15 May 2014

Keywords:

Face hallucination
Sparse representation
Patch size
Regularization parameter

ABSTRACT

Owing to the excellent ability to characterize the sparsity of natural images, ℓ_1 norm sparse representation is widely applied to face hallucination. However, the determination on two key parameters such as patch size and regularization parameter has not been satisfactorily resolved yet. To this end, we proposed a novel parameter estimation method to identify them in an analytical way. In particular, the optimal patch size is derived from the sufficient condition for reliable sparse signal recovery established in compressive sensing theory. Furthermore, by interpreting ℓ_1 norm SR as the corresponding maximum *a posteriori* estimator with Laplace prior constraint, we obtain an explicit expression for regularization parameter in statistics of reconstruction errors and coefficients. Our proposed method can significantly reduce the computational cost of parameter determination while without sacrificing numerical precision and eventual face hallucination performance. Experimental results on degraded images in simulation and real-world scenarios validate its effectiveness.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Face super-resolution, or face hallucination, refers to the technique of estimating a high-resolution (HR) face image from low-resolution (LR) face image sequences or a single LR one. Due to restricted imaging conditions in many scenarios, it is hard to capture HR face images, and thus face hallucination is extensively used for pre- and/or post-processing in video applications, such as video surveillance and face recognition. In their pioneering work on face hallucination [1], Baker and Kanade employed a Bayesian approach to infer the missing high-frequency components of an input LR image from a parent structure with HR/LR training samples, leading to a large magnification factor with relatively good results. Liu et al. [2] proposed a two-step statistical modeling approach that integrates global structure reconstruction with local detail refinement.

Following [1,2], learning-based face hallucination approaches have gained great popularity in recent years. The main idea is to estimate an HR face image from a single LR face image with the help of a training set of HR and LR image pairs. For example, singular value decomposition [3] and morphological component analysis [4] are respectively used to learn mapping coefficients from LR–HR training pairs. Owing to the excellent performance grasping salient properties of natural images, sparse representation (SR) has attracted more attention in face hallucination. Yang et al. [5]

are the first to introduce SR to image super-resolution, where images are approximated by an over-complete dictionary for adaptive sparse image decompositions. This work spurred much follow-up research on SR based face hallucination. In [6], authors presented a dual-dictionary learning method to recover more image details, in which not only main dictionary but also residual dictionary are learned by sparse representation.

In addition to the generic sparsity prior, some specific image priors are further exploited to boost performance in SR based image restoration. Dong et al. [7] proposed non-locally centralized sparse representation to explore the image nonlocal self-similarity. For a class of highly structured objects, such as human faces, the prior of facial positions is of importance and can be utilized to retain the holistic structure of face images. Following this idea, Ma et al. [8] proposed a sparse representation and position prior based face hallucination method. This method classifies high- and low-resolution atoms to form local dictionaries according to the different regions of human face and then uses different local dictionaries to hallucinate the corresponding regions of a face. Generally, SR based methods may select very distinct patches that are far from the input patch to favor sparsity and consequently result in dissimilarity in terms of Euclidean distance. To address this problem, literature [9] introduced a similarity constraint into sparse representation to promote accuracy and stability.

Sparse representation is effective in face hallucination problem when sufficient observations are available, but there are at least two questions we need to answer. As usually done, face hallucination is performed on the basis of small image patches instead of

* Corresponding author.

E-mail address: wzy_hope@163.com (Z. Wang).

a whole image. Thus, firstly, how to choose the optimal patch size given an observed image is worth investigating. Secondly, SR uses a regularization parameter λ to reach a reasonable balance between regression error and sparsity penalty, yet how to efficiently determine λ is unresolved.

To the best of our knowledge, very few face hallucination literatures address patch size choice in a reasoning way. Wang [10] presented an algorithm to detect the lower bound of patch size by measuring the range of the local features in the texture. Kwatra et al. [11] proposed a graph-cut based method, in which the whole input texture is used as a patch and is cut when stitching to other patches. However, these two methods are primarily proposed for texture synthesis in pixel domain (e.g., image in-painting) rather than learning-based super-resolution.

In contrast, classic parameter identification techniques have ever been used to find regularization parameter in super-resolution, such as generalized cross validation (GCV) [12] and L-curve [13]. L-curve is a tool for showing the parametric plot of the error versus penalty with the regularization parameter λ varied. The value of λ at the corner of the L-curve is the desired outcome. GCV produces asymptotically optimum value by iteratively solving a constrained minimization problem. GCV and L-curve can provide high quality solutions, whereas they are computationally expensive.

In practice, representative learning-based face hallucination approaches [4–9] have to use empirical parameter for patch size or λ . All patch-based methods mentioned above set an empirical value for patch size without any theoretical justification. Similarly, the choice of the regularization parameter is based on trial-and-error experiments, which is actually a manual tuning procedure. Although this empirical practice is relatively reliable, it is cumbersome and computationally intensive because of repeated manual trials.

In this paper, we take advantage of the well-known results from the compressive sensing theory to deduce the optimal patch size. As stated by compressive sensing theory, to faithfully recover sparse signal, the dimensions of observed and sparse signals as well as the sparsity should satisfy a pre-known constraint relationship. For SR based face hallucination, the dimensions of observed and sparse signals correspond to the patch size and the number of training images, respectively. Hence the proper patch size may be derived from this well-established constraint equation. Additionally, ℓ_1 norm SR can be interpreted as maximum *a posteriori* (MAP) estimator with Laplace prior imposed on solution. Under MAP framework, regularization parameter λ completely depends on the statistics of noise and coefficients, which actually implies an explicit way for determining λ . In our proposed method, the patch size and regularization parameter are analytically tractable, leading to high efficiency as well as pretty convenience. Experimental results in face hallucination task validate its effectiveness.

The remainder of this paper is organized as follows. Section 2 presents the proposed parameter estimation method in detail. Experimental results are shown in Section 3. In Section 4, we conclude the paper.

2. Proposed method

In this section, we particularly address the problem in determining patch size and regularization parameter. Fig. 1 outlines the patch-wise face hallucination framework, consisting of four parts. Among them, the training of SR coefficients and the reconstruction of HR images are based on the SR model introduced in Section 2.1, whose specific implementation details are referred to literatures [8,9]. In this paper, we mainly focus on the determination technique of patch size and regularization parameter λ .

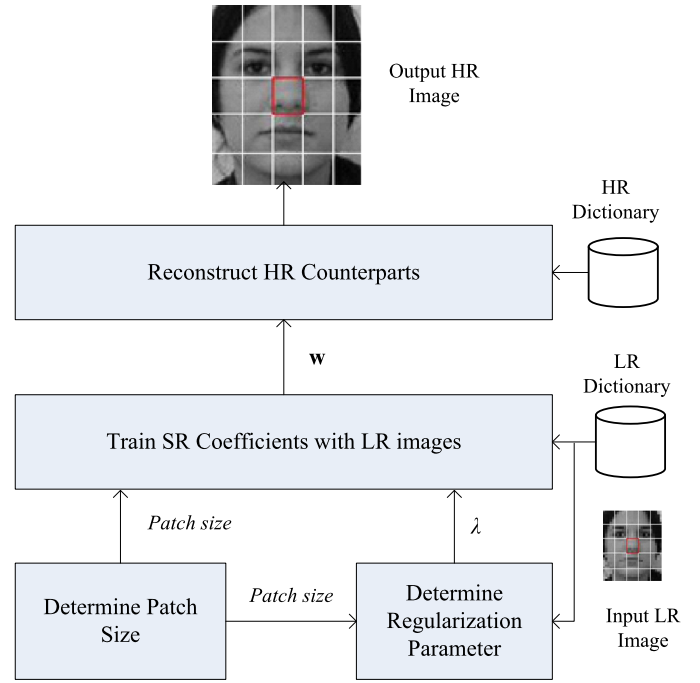


Fig. 1. Outline of the patch-wise face hallucination framework.

2.1. Preliminaries

In dictionary-learning-based face hallucination methods, image patches are represented as a linear combination of elements from an appropriately chosen over-complete dictionary, namely, $\mathbf{x} = \mathbf{Y}\mathbf{w}$. The linear combination coefficients denoted by \mathbf{w} can be learned with the observed LR image over dictionary \mathbf{Y} . Particularly, for SR based face hallucination, the optimal coefficients are obtained by solving ℓ_1 minimization problem. To be more precise, let N be the dimension of an input sample (usually an image patch with size $\sqrt{N} \times \sqrt{N}$) and M be the number of basis samples in training set, for a given input sample $\mathbf{x} \in \mathbb{R}^{N \times 1}$ and a training set $\mathbf{Y} \in \mathbb{R}^{N \times M}$ with each column being an individual training sample, face hallucination via SR can be typically formulated in the following Lagrangian form:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \{ \|\mathbf{x} - \mathbf{Y}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1 \} \quad (1)$$

where $\mathbf{w} \in \mathbb{R}^{M \times 1}$ is an unknown coefficient vector, whose entries w_m , $m = 1, 2, \dots, M$ are associated with all bases in training set. $\lambda \geq 0$ is an appropriately chosen regularization parameter, controlling the tradeoff between the reconstruction error and the ℓ_1 norm penalty. The optimal coefficient vector \mathbf{w}^* can be readily obtained by solving an SR problem.

The primary task of face hallucination is to reconstruct the HR face image from the input LR face image with the help of a training set composed of HR and LR image patches. Optimal coefficients are trained with the input LR image over LR dictionary. According to the manifold similarity paradigm in LR and HR spaces [14], the learning algorithm then maps the local geometry of LR patch space to an HR one, generating HR patch as a linear combination of HR basis patches. Regularization parameter λ and patch size should be predetermined prior to formal face hallucination. In the next subsections, we will discuss the estimate methods on them.

2.2. Patch size

Patch dimension (or patch size) is a key parameter in patch-based face hallucination. It indicates the number of pixels in a

patch and has a considerable impact on the final hallucination results. If the patch size is very small, the co-occurrence prior in LR and HR spaces is too weak to make the prediction meaningful. To the contrary, if the patch size is too large, not only a huge training set is needed to find proximity patches for the current observations, but also the super-resolved face images may be smoothed with some visual details lost. Under given training dictionary, previous practice shows that only a specific value of patch size can result in the best performance [2,4–9]. For example, a patch size of 4×4 pixels is the most favorable one in dictionary learning with 360 basis images. Does this happen by chance? In the following, we justify the choice of patch size from the prospective of compressive sensing theory.

Compressive sensing theory states that minimizing ℓ_1 norm can stably recover the sparse signal from in-complete and inaccurate measurements. Let \mathbf{x} be M -dimensional unknown vector with K -largest non-zero entries, and let $\mathbf{y} = \mathbf{Ax} + \boldsymbol{\epsilon}$ be an observed vector in R^N , where $\mathbf{A} \in R^{N \times M}$ is the known measurement matrix and $\boldsymbol{\epsilon}$ denotes additive noise. It is possible to reliably recover \mathbf{x} from \mathbf{y} using very few measurements (i.e., $N \ll M$) when the RIP (restricted isometry property) condition is satisfied. For Gaussian or Bernoulli random matrices, Candès et al. [15] and Donoho [16] derived RIP of order K as

$$K \approx N / \log(M/N) \quad (2)$$

Nevertheless, face hallucination involves deterministic matrices constructed by training images rather than completely random matrices. According to the result $K \approx \sqrt{N} \log N / \log(M/N)$ by R.A. DeVore [17] on polynomial matrices, a larger N is allowed for a given M when deterministic matrices are concerned. For the sake of reliable inverse recovery of image patches, we make use of Eq. (2) to obtain an approximate but a smaller patch size.

For a known sparsity index K , Eq. (2) can be seen as a transcendental equation with respect to N . This fact motivates us to deduce patch size N from a given K by finding a solution to Eq. (2). Therefore, the sparsity index K of unknown vector \mathbf{x} should be figured out in advance. Literatures [18,19] showed that the sparse solution having K non-zeros found by ℓ_1 minimization is loosely bounded by

$$K \leq \lfloor (N+1)/3 \rfloor \quad (3)$$

Substituting Eq. (3) into Eq. (2), we then have

$$\lfloor (N+1)/3 \rfloor \approx N / \log(M/N) \quad (4)$$

Alternatively, the patch size N can be solved in Eq. (4). More specifically, by a reasonable approximation with $\lfloor (N+1)/3 \rfloor \approx N/3$, the solution to Eq. (4) can be easily obtained: $N \approx M / \exp(3)$. Because image patches are usually shaped in square, we further convert it into an approximate squared integer by

$$N \approx \lfloor \sqrt{M / \exp(3)} \rfloor^2 \quad (5)$$

Eq. (5) gives the analytical solution for the most optimal patch size under a given number of training images. Take the training set having $M = 360$ bases as a concrete example, we acquire a solution of $N \approx 16$. Hence the optimal patch size will be chosen as 4×4 pixels. In contrast to empirical practice in manual tuning, this method makes us free from trivial trials gracefully.

2.3. Regularization parameter

The regularization parameter λ has a significant influence on performance of the underlying representation methods. How to choose it is crucially important to achieve good reconstruction

performance. The optimum setting enables an appropriate trade-off between reconstruction accuracy and robustness against noise. Otherwise, too large values will lead to excessive smoothness (or low-precision prediction) while too small values will result in poor robustness against noise on the contrary [20]. In practice, previous face hallucination methods [4–9,20] empirically specify λ by keeping manual tuning until the best ultimate results are achieved. This empirical practice is relatively reliable, but it is cumbersome and computationally intensive suffering from repeated manual probing. Therefore, we intend to develop an automatic estimation method based on MAP framework.

To deduce the analytical expression of λ in Eq. (1), we briefly present the MAP inference on ℓ_1 regularization. Under MAP formulation [21], the optimal coefficients are estimated by the following Bayesian equation:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \{ \log P(\mathbf{w}) + \log P(\mathbf{x}|\mathbf{w}) \} \quad (6)$$

where $P(\mathbf{x}|\mathbf{w})$ and $P(\mathbf{w})$ denote the conditional probability and prior probability, respectively. Generally, noise is assumed to obey zero-mean i.i.d. Gaussian:

$$P(\mathbf{x}|\mathbf{w}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{Yw}\|_2^2\right) \quad (7)$$

where σ^2 describes the noise level.

As usually stated, the coefficient vector $\mathbf{w} = [w_1, w_2, \dots, w_M]^T$ is governed by a zero-mean i.i.d. multivariate Laplace distribution, namely,

$$P(\mathbf{w}) = \frac{1}{(2\mu)^M} \exp\left(-\frac{\|\mathbf{w}\|_1}{\mu}\right) \quad (8)$$

where $\mu = \frac{\sigma_w}{\sqrt{2}}$ is a scale parameter indicating the diversity, and σ_w describes the standard variance of coefficients.

Substitution of Eqs. (7) and (8) into Eq. (6) leads to

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \left\{ N \log \frac{1}{\sqrt{2\pi}\sigma} + M \log \frac{1}{2\mu} - \frac{1}{2\sigma^2} \left\{ \|\mathbf{x} - \mathbf{Yw}\|_2^2 + \frac{2\sigma^2}{\mu} \|\mathbf{w}\|_1 \right\} \right\} \quad (9)$$

Because M , N , μ and σ are fixed constants, the objective of Eq. (9) is actually equivalent to Eq. (1) with $\lambda = \frac{2\sigma^2}{\mu}$. Replacing μ with $\mu = \frac{\sigma_w}{\sqrt{2}}$, we get the formal expression of regularization parameter in the following form:

$$\lambda = \frac{2\sqrt{2}\sigma^2}{\sigma_w} \quad (10)$$

σ and σ_w represent the standard deviations of noise and coefficients, respectively.

Eq. (10) specifies λ in an analytical way, yet it contains statistics on noise and coefficients, denoted by unknown parameters σ and σ_w . Exploiting the statistics of coefficients actually requires a large number of statistical samples. However, λ is yet unknown, we cannot generate these samples using ℓ_1 norm regularization method. Alternatively, we can obtain a rough estimate of coefficients using least squares (LS) (because it does not include regularization term). On the basis of obtained least-square coefficients, we then reconstruct the observed patch so as to predict the reconstruction errors. When sufficient statistical samples of coefficients and errors are available, the parameters σ and σ_w can be estimated with maximum likelihood (ML) method.

Fig. 2 shows the statistical distributions of the samples of coefficients and reconstruction errors, where actual distributions are

generated with histogram method. Laplace or Gaussian fitting to their actual histograms are obtained with ML method. It is clear that coefficients statistically obey Laplace distribution, while Gaussian fitting approximates the distribution of reconstruction errors. Since the statistics of coefficients and errors can be reliably modeled by Laplace and Gaussian distributions, the sufficient estimation precision of regularization parameter will be expected. Admittedly, this automatic method involves an extra least-square analysis to prepare for statistics. The entire steps for estimating regularization parameter are summarized in Algorithm 1.

Algorithm 1 Regularization parameter estimation.

1. **Input:** LR training set \mathbf{Y} , an LR image \mathbf{x} , and *patch_size*.
 2. Divide each of the training images and the input LR image into a certain number of small patches according to *patch_size*.
 3. **For** each patch located at (i, j) in \mathbf{x} **do**
 - Compute the least-square coefficient vector $\mathbf{w}^*(i, j)$ for the input LR image patch $\mathbf{x}(i, j)$ over the LR training image patches $\mathbf{Y}(i, j)$:

$$\mathbf{w}^*(i, j) = \arg \min_{\mathbf{w}(i, j)} \{ \|\mathbf{x}(i, j) - \mathbf{Y}(i, j)\mathbf{w}(i, j)\|_2^2 \}$$
 - Synthesize the LR patch and compute reconstructed error $\mathbf{e}(i, j)$:

$$\mathbf{e}(i, j) = \mathbf{x}(i, j) - \mathbf{Y}(i, j)\mathbf{w}^*(i, j)$$
 - Collect coefficient samples and error samples by concatenating those of each patch.
 4. **End for**
 5. Based on the collected coefficient and error samples, compute respective standard variances σ_w and σ_e .
 6. Compute regularization parameter λ with Eq. (10).
 7. **Output:** regularization parameter λ .
-

3. Experiments and results

In this section, we conduct face hallucination experiments to verify the proposed method. We first briefly describe the experimental data and method, and then show the experimental results on patch size and regularization parameter.

3.1. Experimental dataset and method

Experiments are performed on two public face databases: FEI face database [22] and CAS-PEAL-R1 face database [23]. The former contains 400 images from 200 subjects (100 men and 100 women). Among them, 360 images are randomly chosen as the training set, and the remaining 40 are used for testing. CAS-PEAL-R1 database contains 30871 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. In both datasets, all the test subjects are completely absent in the training set. The HR images are cropped to 120×100 pixels, and the simulated LR images with 30×25 pixels are generated by smoothing and down-sampling with a factor of 4.

For given patch size and λ , we carry out face hallucination applying the position–patch scheme in [8,9]. According to patch size, we firstly divide the input LR face image into a certain number of mutually overlapping small patches. Overlapping region is used to smooth blocking artifacts suffering from patch-wise manner. The optimal linear coefficients of each patch are resolved with Eq. (1). By keeping the coefficients and replacing the LR dictionary with the corresponding HR one, a new HR patch of the same position can be linearly synthesized. Finally, by concatenating all the HR patches to their corresponding positions and averaging pixel values in the overlapping regions, we can get an estimate of the HR

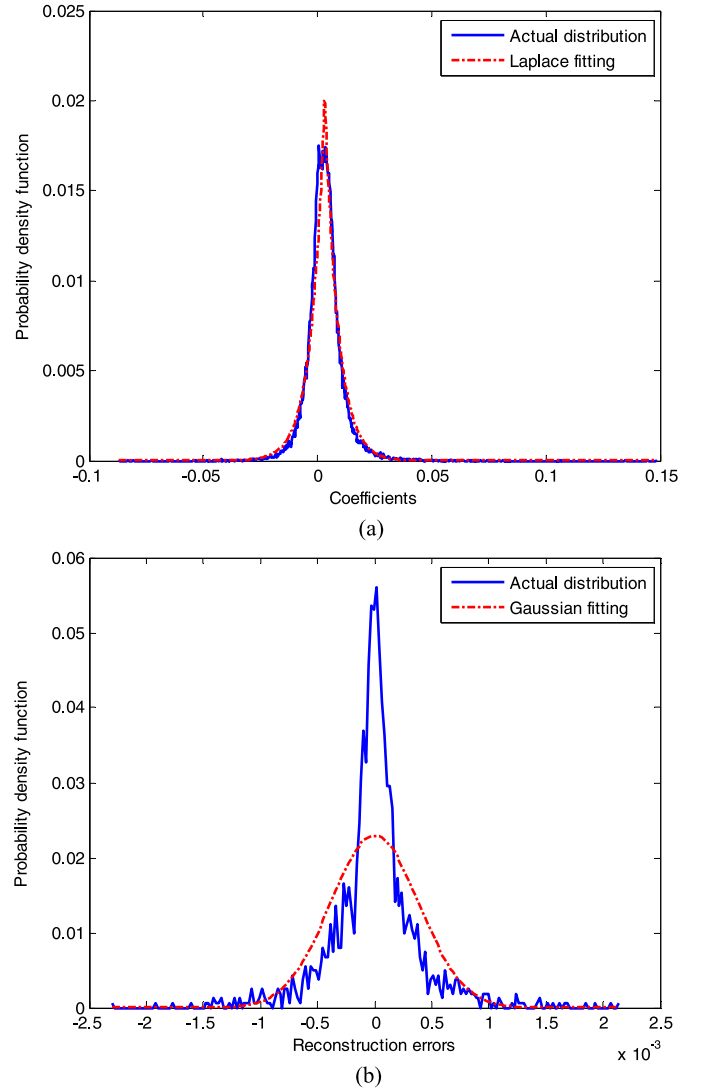


Fig. 2. Statistics of coefficients and reconstruction errors: (a) coefficient samples; (b) reconstruction error samples.

face. The face hallucination performance under varied patch sizes and λ values is evaluated by objective metrics, i.e., PSNR and structural similarity (SSIM) index.

3.2. Experiments on patch size

As discussed in Section 2.2, the optimal patch size can be deduced for sparse representation based face hallucination. These experiments aim to validate the applicability of theoretical results under different number of basis images. We firstly carry out an experiment using FEI face database with the number of training basis being 360. Face hallucination tasks are executed with respect to a set of patch sizes, including 2×2 , 3×3 , 4×4 , 5×5 , 6×6 . We then compute PSNR as well as SSIM for respective hallucinated images. As experimental results shown in Fig. 3, the patch size of 4×4 pixels indeed gives the highest PSNR and SSIM values. This completely agrees with the computed value of 16 pixels in Section 2.2, which shows that our proposed method can determine patch size faithfully.

In another experiment on CAS-PEAL-R1 database, the larger number of training basis (i.e., 1000) is tested. The experimental results with respect to patch sizes of 5×5 , 6×6 , 7×7 , 8×8 , 9×9 are shown in Fig. 4. According to Eq. (5), the optimal patch

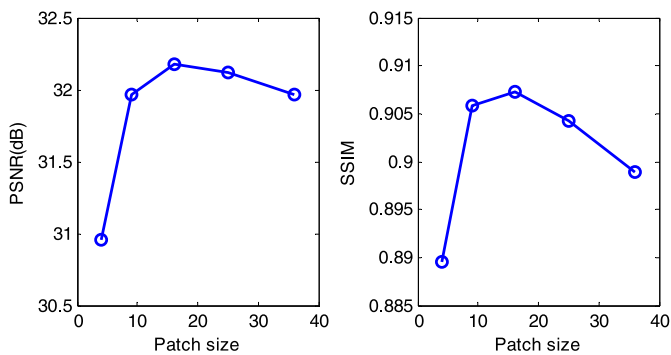


Fig. 3. The average PSNR and SSIM versus different patch sizes in FEI database.

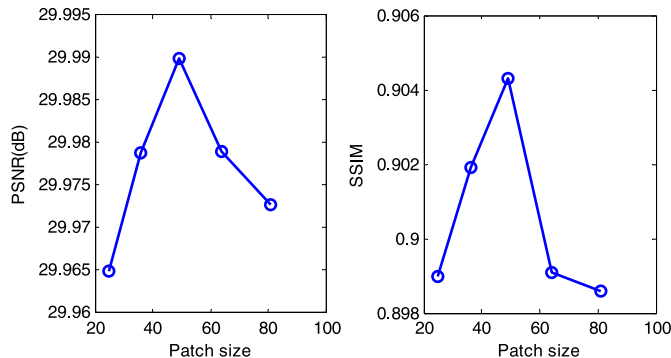


Fig. 4. The average PSNR and SSIM versus different patch sizes in CAS-PEAL-R1 database.

size should be 7×7 , which just corresponds to the peaks of PSNR and SSIM in Fig. 4.

Since the patch size of 4×4 pixels is verified to offer the best results for FEI database involving 360 basis images, we will use this value in the subsequent experiments. Specifically, we set the LR patch size as 4×4 pixels and the overlap between neighborhood patches as 1 pixel, while the corresponding HR patch size is 16×16 pixels with an overlap of 4 pixels (because of 16-times magnification). Similarly, for the experiments involving 1000 basis images in CAS-PEAL-R1 database, the LR patch size of 7×7 pixels and the HR patch size of 28×28 pixels are put into practice.

3.3. Experiments on regularization parameter

To compare the accuracy and computational cost in determining regularization parameter λ , we choose the commonly used empirical method in literatures [5,8,9] and well-known L-curve method [13] as anchors (because GCV [12] is obtained for ridge regression rather than SR). Three classes of images are used in this experiment, including noise free images, simulated noisy images and real-world surveillance images. The original noiseless images are randomly selected from FEI face database. We add zero-mean Gaussian white noise (the noise levels $\sigma = 5$ and $\sigma = 10$) to original images to get simulated noisy images. The real-world surveillance images are captured by commercial surveillance camera in a low lighting and at a distanced location and thus unavoidably contain noise and blurring effects. The statistics of experimental results on FEI database are shown in Table 1 and Table 2, and some of randomly selected subjective results are shown in Fig. 5 and Fig. 6.

In Table 1, we not only show the values of λ by different methods but also report the objective metrics with respective λ . Note that, the objective metrics are not applicable for real-world case since the reference HR images are non-existent. It is clear that the estimated values of λ by our method closely approach those of

Table 1
Comparison of regularization parameter λ by different methods in FEI database.

Images	Methods	λ	PSNR	SSIM
Noise free	Proposed	$7.7e-5$	31.68	0.9077
	L-curve	$9.6e-5$	31.69	0.9078
	Empirical	$1.0e-4$	31.70	0.9080
$\sigma = 5$ Gaussian	Proposed	$1.0e-2$	29.26	0.8433
	L-curve	$6.7e-3$	29.45	0.8461
	Empirical	$5.0e-3$	29.52	0.8460
$\sigma = 10$ Gaussian	Proposed	$2.2e-2$	27.31	0.7842
	L-curve	$2.1e-2$	27.33	0.7844
	Empirical	$2.0e-2$	27.36	0.7842
Real-world	Proposed	$2.1e-2$	n/a	
	L-curve	$1.7e-2$		
	Empirical	$1.0e-2$		

Table 2
Comparison of regularization parameter λ by different methods in CAS-PEAL-R1 database.

Images	Methods	λ	PSNR	SSIM
Noise free	Proposed	$2.7e-4$	29.24	0.8966
	L-curve	$2.9e-4$	29.26	0.8967
	Empirical	$3.0e-4$	29.27	0.8968
$\sigma = 5$ Gaussian	Proposed	$3.1e-2$	27.43	0.8332
	L-curve	$2.5e-2$	27.61	0.8349
	Empirical	$2.0e-2$	27.66	0.8350
$\sigma = 10$ Gaussian	Proposed	$7.2e-2$	26.36	0.7802
	L-curve	$6.8e-2$	26.40	0.7807
	Empirical	$6.0e-2$	26.42	0.7805

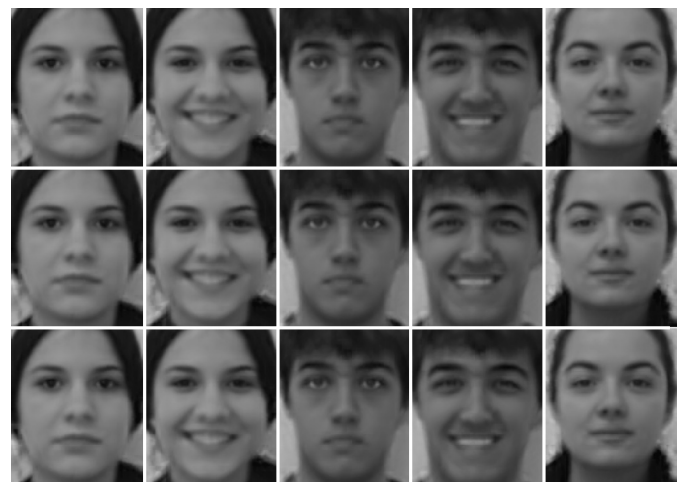


Fig. 5. Results by different methods for simulated LR faces in FEI database. Top: proposed method; middle: L-curve method; bottom: empirical method.

empirically tuning and L-curve in all cases. Irrespective of the real-world case, the hallucinated results among three methods almost make no differences in terms of average PSNR and SSIM. Fig. 6 also shows that there are no visible distinctions among the subjective results by different methods under real-world application scenarios.

In addition to performing experiments on the FEI face database, we also conduct experiments on CAS-PEAL-R1 database. Provided that the objective performance in terms of PSNR and SSIM cannot be evaluated in the scenario of real-world surveillance images, we only testify the simulated LR images. In a similar experimental method to FEI database, the simulated LR images are super-resolved by 16 times and the statistics of experimental results are tabulated in Table 2. Some of randomly selected visual results



Fig. 6. Results by different methods for real-world LR faces in FEI database. Top: proposed method; middle: L-curve method; bottom: empirical method.

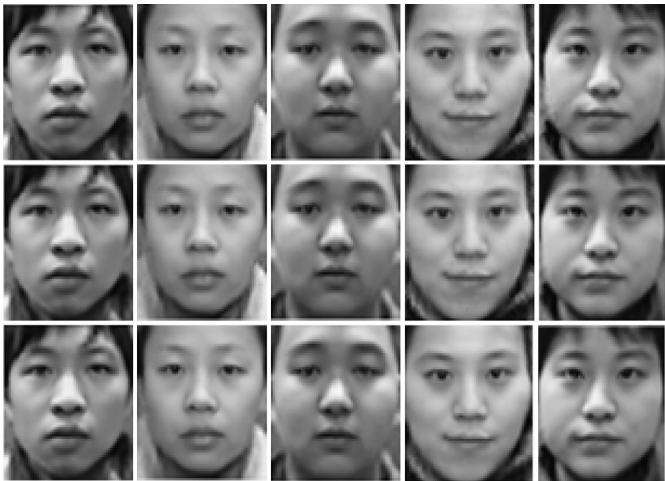


Fig. 7. Results by different methods for simulated LR faces in CAS-PEAL-R1 database. Top: proposed method; middle: L-curve method; bottom: empirical method.

from the noise free test are depicted in Fig. 7. Again, we can also draw the same conclusion that our proposed method maintains the comparable numerical accuracy for regularization parameter λ and hallucination performance as well. Meanwhile, we also find that the results in Table 2 differ from those in Table 1 considerably. This may be attributed to inherent differences of image statistics in FEI face database and CAS-PEAL-R1 face database. On one hand, face images in FEI have been pre-aligned by provider while those in CAS-PEAL-R1 are aligned by ourselves with five manually selected feature points. On the other hand, as samples demonstrated in Fig. 5 and Fig. 7, the face images in FEI exhibit more uniform facial poses than ones in CAS-PEAL-R1. Obviously, these factors will result in variations in hallucination effects.

As for computational cost, our method only employs an extra least-square estimate to prepare for statistical samples, while L-curve and empirical methods require a larger number of computations due to complex iterative trials. Fig. 8 illustrates an example on the iterative procedure in L-curve method [13] using FEI database, where it takes 7 iterations to achieve convergence with the optimal value of λ occurring at the corner of the L-curve. In the empirical method, the candidates of λ are altered in a heuristic manner based on trial-and-error tactics and the one corresponding to the maximum PSNR or the best visual effect will be selected. Fig. 9 shows the PSNR curve with respect to varied λ in a typical

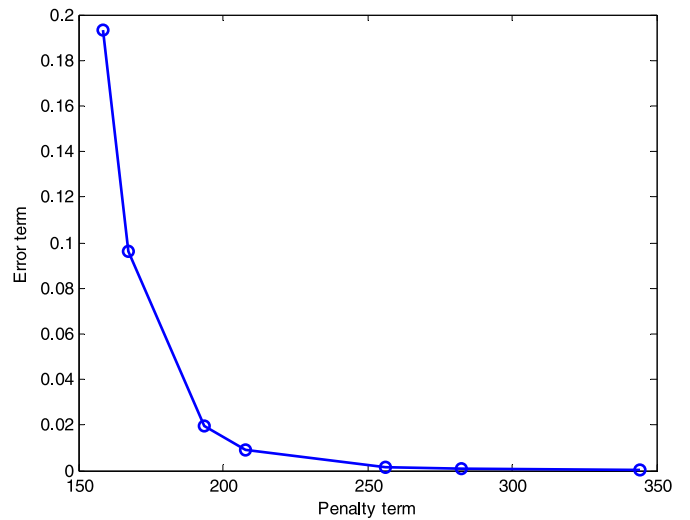


Fig. 8. The iterative process in L-curve method.

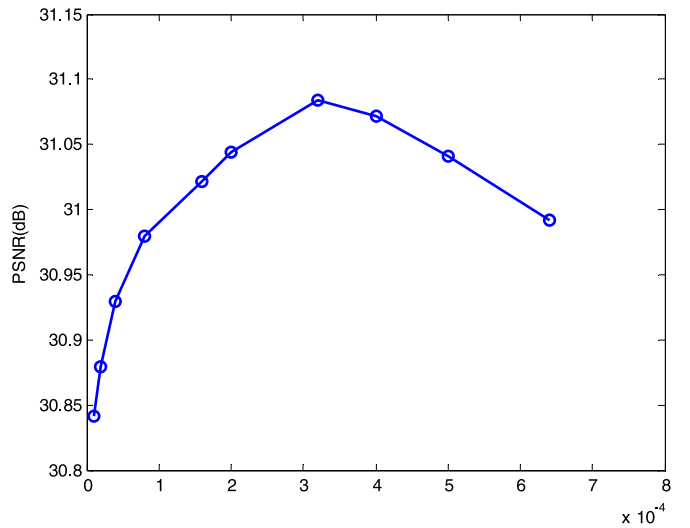


Fig. 9. The average PSNR versus λ in the empirical method.

Table 3

Comparison of computational complexity by different methods.

Methods	Proposed	L-curve	Empirical
Number of analysis	1 (LS)	7 (SR)	11 (SR)
Number of synthesis	1 (LR)	7 (LR)	11 (HR)

example on the empirical method, where 11 times of probing are spent to find an optimal value.

We particularly investigate the computational complexity of different methods. To simplify evaluation, we roughly count the number of involved analysis and synthesis tasks in determining λ while ignoring some other relatively minor operations, such as computing reconstruction error and locating the λ position of the L-curve. Specifically, analysis task is responsible for solving the ℓ_1 norm SR or ordinary LS, and synthesis task deals with reconstructing LR or HR images. As shown in Table 3, our method only needs LS analysis and LR synthesis, while the other two involve optimization on more complex SR problem. In order to examine PSNR or subjective quality on hallucinated results, empirical method has to reconstruct HR images, leading to higher computational cost. Regardless of specific optimization problem and image resolution, our method involves the least computing tasks (only 1 times), while L-curve and empirical methods require average 7 and 11 times,

respectively. Evidently, our method is much more computationally efficiency than counterparts.

4. Conclusions

In this paper, we have proposed a novel method to automatically estimate the patch size and regularization parameter in SR based face hallucination. Different from empirical practice or L-curve method, our method enjoys both advantages of high computational efficiency and pretty convenience, while without sacrificing tuning precision. Experimental results on FEI and CAS-PEAL-R1 face databases as well as real-world images have confirmed its applicability in terms of numerical accuracy, computational cost and face hallucination performance. The proposed parameter estimation method can be readily incorporated into many existing SR based super-resolution approaches. Moreover, this reasoning paradigm can be generalized to similar patch-based super-resolution techniques employing other kinds of regularization, e.g., ridge regression.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (61070080, 61170023, 61172173, 61231015, 61303114, 61172174, 6130211), the Major National Science and Technology Special Projects (2010ZX03004-003-03, 2012YQ16018505), and the Fundamental Research Funds for the Central Universities (2042014kf0286, 2042014kf0212, 2042014kf0025, 2042014kf0250).

References

- [1] S. Baker, T. Kanade, Limits on super-resolution and how to break them, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (9) (2002) 1167–1183.
- [2] C. Liu, H. Shum, W. Freeman, Face hallucination: theory and practice, *Int. J. Comput. Vis.* 75 (1) (2007) 115–134.
- [3] M. Jian, K.-M. Lam, J. Dong, A novel face-hallucination scheme based on singular value decomposition, *Pattern Recognit.* 46 (2013) 3091–3102.
- [4] Y. Liang, X. Xie, J.-H. Lai, Face hallucination based on morphological component analysis, *Signal Process.* 93 (2013) 445–458.
- [5] J. Yang, H. Tang, Y. Ma, T. Huang, Image super-resolution via sparse representation, *IEEE Trans. Image Process.* 19 (11) (2010) 2861–2873.
- [6] Jian Zhang, Chen Zhao, Ruiqin Xiong, Siwei Ma, Debin Zhao, Image super-resolution via dual-dictionary learning and sparse representation, in: *IEEE Int. Symposium on Circuits and Systems*, 2012, pp. 1688–1691.
- [7] Weisheng Dong, Lei Zhang, Guangming Shi, Xin Li, Nonlocally centralized sparse representation for image restoration, *IEEE Trans. Image Process.* 22 (4) (2013) 1620–1630.
- [8] X. Ma, H.Q. Luong, W. Philips, H. Song, H. Cui, Sparse representation and position prior based face hallucination upon classified over-complete dictionaries, *Signal Process.* 92 (2012) 2066–2074.
- [9] Z. Wang, J. Jiang, Z. Xiong, R. Hu, Z. Shao, Face hallucination via weighted sparse representation, in: *IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, 2013, pp. 2198–2201.
- [10] B.J. Wang, An algorithm for detecting optimized patch size for image quilting, *Technical Report*, Microsoft, 2003.
- [11] V. Kwatra, A. Schödl, I. Essa, G. Turk, A. Bobick, Graphcut textures: image and video synthesis using graph cuts, *ACM Trans. Graph.* 22 (3) (2003) 277–286.
- [12] N. Nguyen, P. Milanfar, G. Golub, A computationally efficient super resolution image reconstruction algorithm, *IEEE Trans. Image Process.* 10 (3) (2001) 573–583.
- [13] N.K. Bose, S. Lertrattanapanich, J. Koo, Advances in super-resolution using L-curve, in: *IEEE Int. Sym. on Circuits and Systems*, 2001, pp. 433–436.
- [14] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (5500) (2000) 2323–2326.
- [15] E.J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurement, *Commun. Pure Appl. Math.* 59 (8) (2006) 1207–1223.
- [16] D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52 (4) (2006) 1289–1306.
- [17] Ronald A. DeVore, Deterministic constructions of compressed sensing matrices, *J. Complex.* 23 (2007) 918–925.
- [18] J. Wright, A.Y. Yang, Arvind Ganesh, S. Shankar Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [19] D. Donoho, Neighborly polytopes and sparse solution of underdetermined linear equations, Technical report: 2005-4, Dept. of Statistics, Stanford Univ., 2005.
- [20] Jiming Li, Weiguo Gong, Weihong Li, Feiyu Pan, Single-image super-resolution reconstruction based on global non-zero gradient penalty and non-local Laplacian sparse coding, *Digit. Signal Process.* 26 (2014) 101–112.
- [21] Richard R. Schultz, Robert L. Stevenson, A Bayesian approach to image expansion for improved definition, *IEEE Trans. Image Process.* 3 (3) (1994) 233–242.
- [22] FEI face database [online], available: <http://fei.edu.br/~cet/face-database.html>.
- [23] W. Gao, B. Cao, S.G. Shan, X.L. Chen, D.L. Zhou, X.H. Zhang, D.B. Zhao, The CAS-PEAL large-scale Chinese face database and baseline evaluations, *IEEE Trans. Syst. Man Cybern., Part A, Syst. Humans* 38 (1) (2008) 149–161.

Zhongyuan Wang received the B.S. degree and M.S. degree in Computer Science from Wuhan University, Wuhan, China, in 1995 and 2001, and he received the Ph.D. degree in Communication and Information System in Wuhan University, in 2008. From 2001, he worked as a member of research staff in National Multimedia Software Engineering Research Center of Wuhan University. His research interests include video compression, multimedia communications.

Ruimin Hu received the B.S. and M.S. degrees from Nanjing University of Posts and Telecommunications, Nanjing, China, in 1984 and in 1990, and Ph.D. degree in Communication and Electronic System from Huazhong University of Science and Technology, Wuhan, China, in 1994. Dr. Hu is the director of National Engineering Research Center for Multimedia Software, Wuhan University and Key Laboratory of Multimedia Network Communication Engineering in Hubei province. He is Executive Chairman of the Audio Video coding Standard (AVS) workgroup of China in Audio Section. He has published two books and over 100 scientific papers. His research interests include audio/video coding and decoding, video surveillance and multimedia data processing.

Zhenfeng Shao obtained his Ph.D. degree on 3D reconstruction of manmade objects from Wuhan University, China, in 2004. He is now professor in State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China. His research interests are Image retrieval, Image fusion, 3D reconstruction and urban remote sensing application.

Zhiqiang Hou received the B.S. and M.S. degrees in Information Science and Engineering from Henan University of Technology, Henan, China, in 2007 and 2010, respectively. He is currently working with the School of Computer for Ph.D. degree in Wuhan University, Wuhan, China. His current research interests include video coding and communications.