

Robust Face Hallucination via Similarity Selection and Representation

Feng Liu^(✉), Ruoxuan Yin, Zongliang Gan, Changhong Chen,
and Guijin Tang

Jiangsu Provincial Key Lab of Image Processing and Image Communication,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China
{liuf, l3010614, ganzl, chenhh, tanggj}@njupt.edu.cn

Abstract. Face image super resolution, also referred to as face hallucination, is aiming to estimate the high-resolution (HR) face image from its low-resolution (LR) version. In this paper, a novel two-layer face hallucination method is proposed. Different from the previous SR methods, by applying global similarity selecting, the proposed approach can narrow the scope of samples and boost the reconstruction speed. And the local similarity representation step make the method have better ability to suppress noise for applications under severe condition. As a general framework, other useful algorithms can also be incorporated into it conveniently. Experiments on commonly used face database demonstrate our scheme has better performance, especially for noise face image.

Keywords: Face hallucination · Eigen face · Global similarity selection · Local similarity representation

1 Introduction

Rapid development of video and image services call for high resolution face images. But existing condition of image and video equipment often can't meet the requirement. Therefore, many single image super-resolution researches, targeting on sensitive regions such as vehicle plates and faces, attract much attention. Face hallucination refers to the technique of reconstructing the latent high-resolution (HR) face from a single low-resolution (LR) face. Recent state-of-the-art face hallucination methods are mostly learning-based. These learning-based methods [1] utilize HR and LR dictionary pair to obtain the similar local geometry between HR and LR samples, and achieve satisfactory result under stationary condition.

Freeman et al. [2] firstly proposed a patch-based Markov network to model the local geometry relationship between LR patches and their HR counterparts which is time consuming. Baker and Kanade [4] employed a Bayesian model to estimate the

This research was supported in part by the National Nature Science Foundation, P. R. China. (No. 61071166, 61172118, 61071091, 61471201), Jiangsu Province Universities Natural Science Research Key Grant Project (No. 13KJA510004), Natural Science Foundation of Jiangsu Province (BK20130867), the Six Kinds Peak Talents Plan Project of Jiangsu Province(2014-DZXX-008), and the "1311" Talent Plan of NUPT.

latent high-frequency components. Inspired by locally linear embedding (LLE), Chang et al. [7] obtained high-resolution image through neighbor embedding (NE) method, but the fixed number of neighbors is unstable and difficult to estimate. Considering that face images have stable structure, Ma et al. [8] took it as prior and proposed a position-patch based method, which only use the patches in the same position. But when the number of training faces is much larger than the LR patch’s dimension, unconstrained least square estimation (LSE) [8, 9] will lead to inaccurate solutions. Yang et al. [12] and Jung et al. [13] introduced sparse representation (SR) method and convex optimization to address this problem, respectively. Although these SR based methods have achieved good results, their performance are strictly limited by the noise level and degradation process. Figure 1 shows the framework of these typical position-patch based methods. To better adapt to real situation, constrained SR methods are proposed. By affixing similarity constrained factor to optimization coefficients [15, 16], adding illumination compensation for constrained factor [17], and introducing adaptive ℓ -q norm to sparse representation [18], these constrained SR methods improved the ability to resist noise. But the computational load of the commonly used SLEP toolbox [23] which solves the optimization problem is significant.

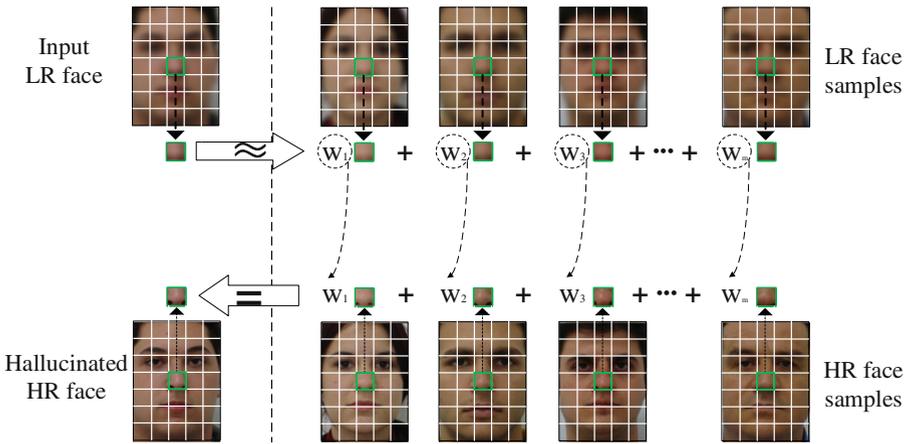


Fig. 1. Outline of the traditional position-patch based face hallucination framework.

In this paper, we prove that the common face database used to face hallucination is over-complete and not all the samples are necessary. We also establish a novel two-layer framework which has a natural parallel structure to infer HR patch. Inspired by the constrained sparse representation methods, we think part of the samples in training set are very different from the input face, which corresponding to the nearly-zero coefficients during optimization. These faces are less important and can even be directly ignored to improve robustness and also reduce computation. We find that reducing the number of trained samples according to their similarities with the input LR face hardly affect the result but save a lot of time, which proves our

assumption above. So a sample selection strategy is needed. Considering that the samples in database are all well-aligned faces and almost under the same condition, we intuitively choose eigenface to present the sample reducing operation. Finally, by using global similarity selection via eigenface to speed up the reconstruction process, and using local similarity representation between the same position patches and neighbor patches to synthesize HR patch, our proposed similarity selection and representation (SSR) method get better performance when the input image is corrupted by noise and different blur kernels. As a general framework, the global comparison and local representation algorithm can also be replaced by other useful face recognition and representation methods conveniently. The proposed method has the following features.

- When the surrounding (neighbor) patches are considered to improve robustness against face position changing, about ten times of calculation will be increased. Our two-layer parallel framework can boost the reconstruction speed.
- By excluding insignificant or dissimilar samples in the training set before reconstruction, the algorithm is further accelerated with no decrease on the quality.
- Using constrained exponential distance as coefficients can guarantee good structure similarities between patches, and meanwhile solve the over-fitting problem caused by noise in sparse representation, which make the method very robust to noise and blur kernel changing in real scenarios.

2 Existing Position-Patch Based Approaches

Let N be the dimension of a patch (usually a patch with size $\sqrt{N} \times \sqrt{N}$) and M be the number of samples in the training set. Given a test patch $\mathbf{x} \in R^{N \times 1}$, the patches at the same position in LR training set is represented as $\mathbf{Y} \in R^{N \times M}$, with m th column \mathbf{Y}^m ($m = 1, \dots, M$) being the patch in sample m . Then the input patch can be represented by the sample patches as:

$$\mathbf{x} = \mathbf{Y}\mathbf{w} + \mathbf{e} \quad (1)$$

where $\mathbf{w} \in R^{M \times 1}$ is the coefficients vector with entry w_m and \mathbf{e} is the reconstruction error vector.

Obviously, solving the coefficients \mathbf{w} is the key problem in patch-based methods. In [7–9], reconstruction coefficients are solved by constrained least square estimation (LSE) as

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{x} - \mathbf{Y}\mathbf{w}\|_2^2 \text{ s.t. } \sum_{m=1}^M w_m = 1 \quad (2)$$

This least square problem's closed-form solution can be solved with the help of Gram matrix, but it becomes unstable when the number of samples M is much bigger than the patch dimension N . Jung et al. [13] introduce sparse representation into face hallucination and convert (2) into a standard SR problem:

$$\min_{\mathbf{w}} \|\mathbf{w}\|_0 \text{ s.t. } \|\mathbf{x} - \mathbf{Y}\mathbf{w}\|_2^2 \leq \varepsilon \quad (3)$$

where ℓ_0 norm counts the non-zero entries number in \mathbf{w} and ε is error tolerance. Yang et al. [12] and Zhang et al. [14] respectively use squared ℓ_2 norm $\|\mathbf{w}\|_2^2$ and ℓ_1 norm $\|\mathbf{w}\|_1$ to replace ℓ_0 norm, which means the statistics of coefficients are constrained by Gaussian and Laplacian distribution.

Similarity constrained SR methods proposed in [15–18] can be formulated as

$$\begin{aligned} \mathbf{w}^* &= \arg \min_{\mathbf{w}} \left\{ \|\mathbf{x} - \mathbf{Y}\mathbf{w}\|_2^2 + \lambda \|\mathbf{D}\mathbf{w}\|_q \right\}, \\ d_{mm} &= \|g\mathbf{x} - \mathbf{Y}^m\|_2, 1 \leq m \leq M \end{aligned} \quad (4)$$

where $\mathbf{D} \in \mathbb{R}^{M \times M}$ is the diagonal matrix which controls the similarity constrains placed on coefficients. The entries d_{mm} on the main diagonal of \mathbf{D} represent the Euclidean distance with gain factor g . Furthermore, in [15–17], q is set to 1 and 2 respectively. In [18], an adaptively selected ℓ_q norm scheme is introduced to improve its robustness against different conditions.

After the coefficients are obtained, the reconstructed HR test patch \mathbf{x}_H can be represented as

$$\mathbf{x}_H = \mathbf{Y}_H \mathbf{w} \quad (5)$$

with coefficients being directly mapped to the patches of HR samples \mathbf{Y}_H .

3 Proposed Method

The methods introduced in Sect. 2 have shown impressive results for experimental noise free faces. But when the noise level, blur kernel and degradation process change, the performance will drop dramatically. It's mainly due to the under-sparse nature of noise, and the local geometry between the high dimension and the low dimension manifolds are no longer coherent since the degradation process has changed. To overcome this problem, we propose global similarity selection and local similarity representation to improve its robustness.

3.1 Global Similarity Selection

In SR based methods, a test LR patch is represented by a large number of LR sample patches through coefficients with sparse. Therefore, heavy calculation is cost on the coefficients with sparsity even the corrupted face is not sparse. So we think not all the samples in the over-complete face database are necessary. Some faces in training set are very different from the input face, which corresponding to very tiny weights during optimization. These faces are not worth occupying so much calculation because they have very limited impact on the results.

Therefore, for the noise corrupted faces, we no longer look for the representation with sparsity but use similarity representation to represent LR patch directly. In order to exclude the dissimilar samples which corresponding to the nearly-zero coefficients, a similarity comparison strategy is needed. And we choose global similarity selection scheme, instead of local (patch) similarity comparison before the reconstruction. This is mainly because the face databases we used are all same size, well aligned and under same lighting condition, global comparison can be reliable enough and very fast.

We intuitively apply Turk and Pentlad's [10] eigenface method, which projects the test face image into the eigenface space, and selects the most similar M faces according to Euclidean distance. Given the LR face database \mathbf{F} , with m th column being sample \mathbf{F}^m . After being normalized by subtracting its mean value, the covariance matrix \mathbf{C} can be obtained by (6), where M is the number of samples in \mathbf{F} . Then the eigenface \mathbf{P} is easy to compute by singular value decomposition, and the projected database \mathbf{F}' is given in (7). Before reconstruction, the input LR face \mathbf{x} is firstly projected to the eigenface space by (8), and the similar faces can be selected between the samples and \mathbf{x} through Euclidean distance, according to $\mathbf{F}'_m - \mathbf{x}'^2_2$.

$$\mathbf{C} = \frac{1}{M} \mathbf{F}_{norm} \mathbf{F}_{norm}^T, \mathbf{F}_{norm} = \mathbf{F} - \text{mean}(\mathbf{F}). \quad (6)$$

$$\mathbf{F}' = \mathbf{F}_{norm}^T \mathbf{P}. \quad (7)$$

$$\mathbf{x}' = [\mathbf{x} - \text{mean}(\mathbf{F})]^T \mathbf{P}. \quad (8)$$

Results shown in Fig. 3 in Sect. 4 demonstrate our assumption perfectly. This global similarity selection method have saved about half of the traditional method's calculation before reconstruction.

3.2 Local Similarity Representation

After picking out the similar faces we need from the entire training set. To make full use of the information in neighbor patches, we establish a framework with parallel two-layer structure which integrates the information of patches surrounding the test LR patch, as shown in Fig. 2. Instead of using all the sample patches at the same position to estimate the test patch (as in Fig. 1), we change the traditional structure into a two layer mode which every sample outputs a middle-layer HR patch before synthesizing the final HR patch.

In Fig. 2, we can notice that the test LR patch's neighbor patches in training samples are marked with dark lines. Let M be the number of samples we picked out and S (S is 9 in Fig. 2) be the number of neighbor patches we used surrounding the center patch. The test LR patch is still $\mathbf{x} \in \mathbb{R}^{N \times 1}$, and all its neighbor patches in sample m ($m = 1, \dots, M$) is represented as $\mathbf{Y}_m^S \in \mathbb{R}^{N \times S}$ with s th column being \mathbf{Y}_m^s ($s = 1, \dots, S$).

For every sample in the LR training set, a weight vector $\mathbf{w}_m \in \mathbb{R}^{S \times 1}$ which represents the similarity between the test patch and its neighbor patches is obtained, the entries w_m^s are computed as

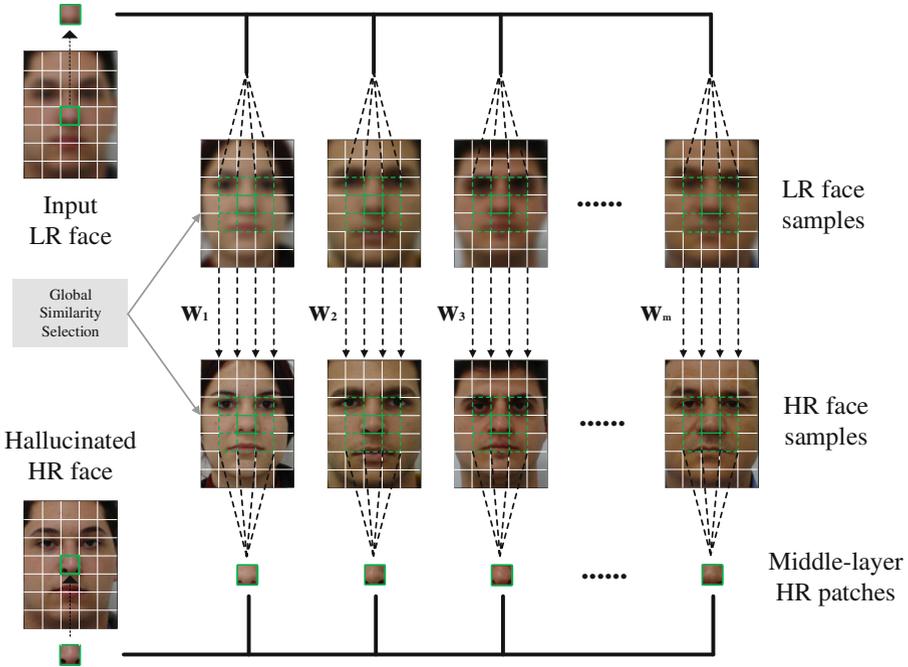


Fig. 2. Proposed parallel two-layer face hallucination framework

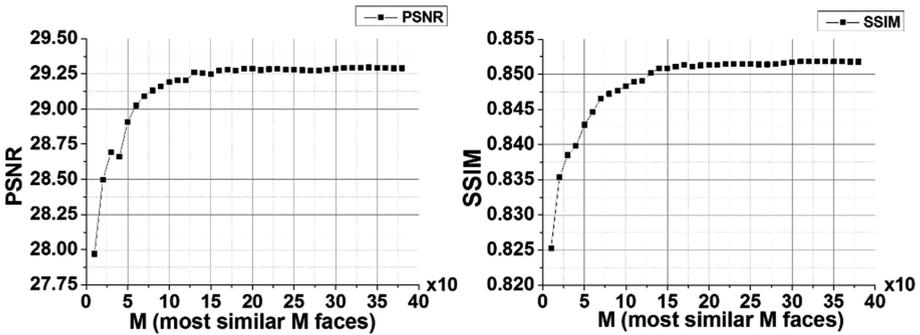


Fig. 3. Average PSNR and SSIM with the number of the similar faces M changes

$$w_m^s = \exp\left(-\left(\|x - Y_m^s\|_2 + \|D(s) - D(x)\|_2\right)\right), 1 \leq s \leq S. \quad (9)$$

The function $D(\cdot)$ calculates the patch distance from the current patch to the center patch. Then the middle-layer HR patch of sample m can be represented as

$$x_{H1m} = \frac{1}{G_m} Y_m^S w_m,$$

$$G_m = \sum_{s=1}^S w_m^s \quad (10)$$

After every sample output its corresponding middle-layer HR patch, the final HR patch can be synthesized simply by

$$\mathbf{x}_{H2} = \frac{1}{\sum_{m=1}^M G_m} \sum_{m=1}^M G_m \mathbf{x}_{H1m}. \quad (11)$$

Finally, by assembling all reconstructed HR patches to the corresponding position and averaging the pixels in overlapped regions. The estimated HR face is obtained. The entire similarity selection and representation face hallucination method is summarized as following steps.

Face Hallucination via SSR

1. **Input:** HR and LR training set with M' samples, LR test face \mathbf{x}_L , number of neighbors S , *patchsize*, *overlap*.
2. Project LR test face into the eigenface space, and pick out the most M similar samples according to Euclidean distance.
3. Divide the input test face and the picked out M training samples into UV patches according to *patchsize* and *overlap*.
4. **For** $i = 1$ to U **do**
5. **For** $j = 1$ to V **do**
6. **For** $m = 1$ to M **do**
 - (a) Calculate similarity weights \mathbf{w}_m between test patch $\mathbf{x}_L(i, j)$ and its neighbors in sample m .

$$w_m^s = \exp(-\|\mathbf{x} - \mathbf{Y}_m^s\|_2 - \|D(\mathbf{x}) - D(s)\|_2), \quad 1 \leq s \leq S$$

- (b) Synthesize middle-layer HR patch over LR neighbor patches \mathbf{Y}_m^S

$$\mathbf{x}_{H1m}(i, j) = \frac{1}{G_m} \mathbf{Y}_m^S \mathbf{w}_m, \quad G_m = \sum_{s=1}^S w_m^s.$$

7. **End for**
8. Synthesize final HR patch at position (i, j)

$$\mathbf{x}_{H2}(i, j) = \frac{1}{\sum_{m=1}^M G_m} \sum_{m=1}^M G_m \mathbf{x}_{H1m}(i, j).$$

9. **End for**
 10. **End for**
 11. Integrate all the reconstructed HR patches to their corresponding position and averaging the pixel values in the overlapped regions.
 12. **Output:** HR hallucinated face \mathbf{x}_H .
-

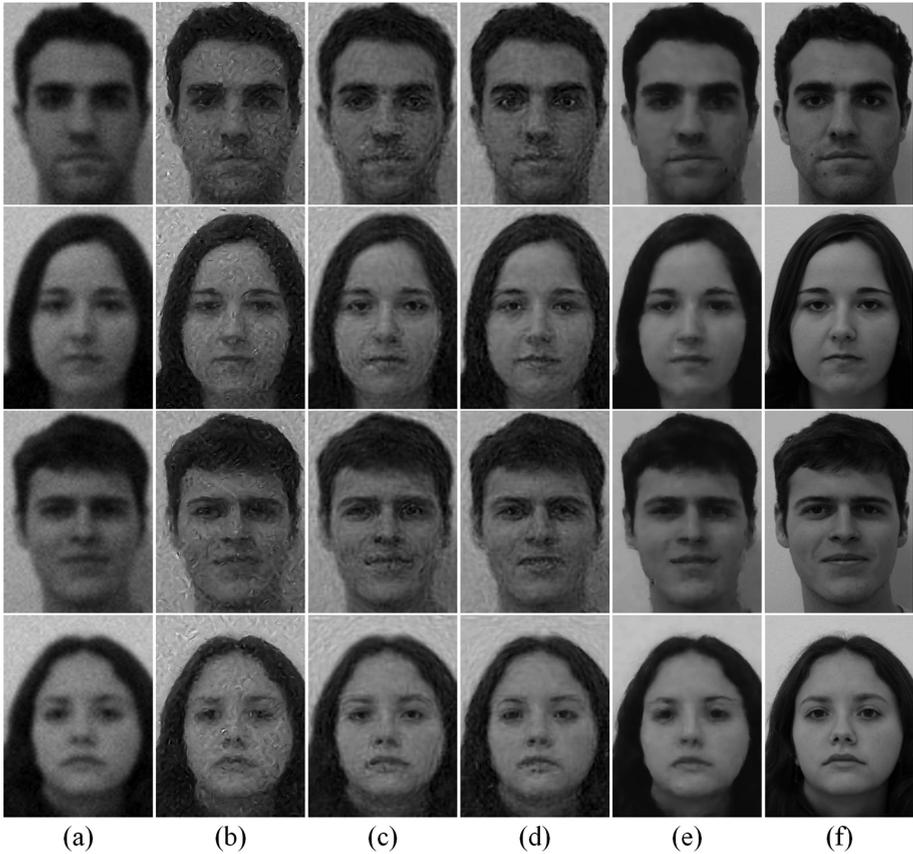


Fig. 4. Comparison of results based on different methods on FEI face database for noisy images ($\sigma = 5$). (a) Bicubic. (b) Chang's NE [7]. (c) Ma's LSR [8]. (d) Jiang's LcR [15]. (e) Proposed SSR. (f) Original HR faces.

4 Experiments and Results

In this section, we conduct face hallucination experiments on the proposed SSR method to testify its performance under inconsistent degradation conditions between the test face image and training face images. The face database we apply is FEI face database [24], which contains 400 faces from 200 adults (100 men and 100 women). Among them, we randomly select 380 faces for training and the rest 20 faces for testing. All the samples are well-aligned and in the same size 360×260 . The LR samples are smoothed and down-sampled with factor of 4. The LR patch size is set to 3×3 pixels with overlap of 1 pixel and the HR patch size is 12×12 pixels with 4 overlapped pixels. The smooth kernel we use in the training phase is fixed by 20×20 Gaussian lowpass filter with standard deviation of 4.

In Fig. 3, we show how the number of similar faces M in the global similarity selection stage affect the quality of reconstruction. We can see that the second half of

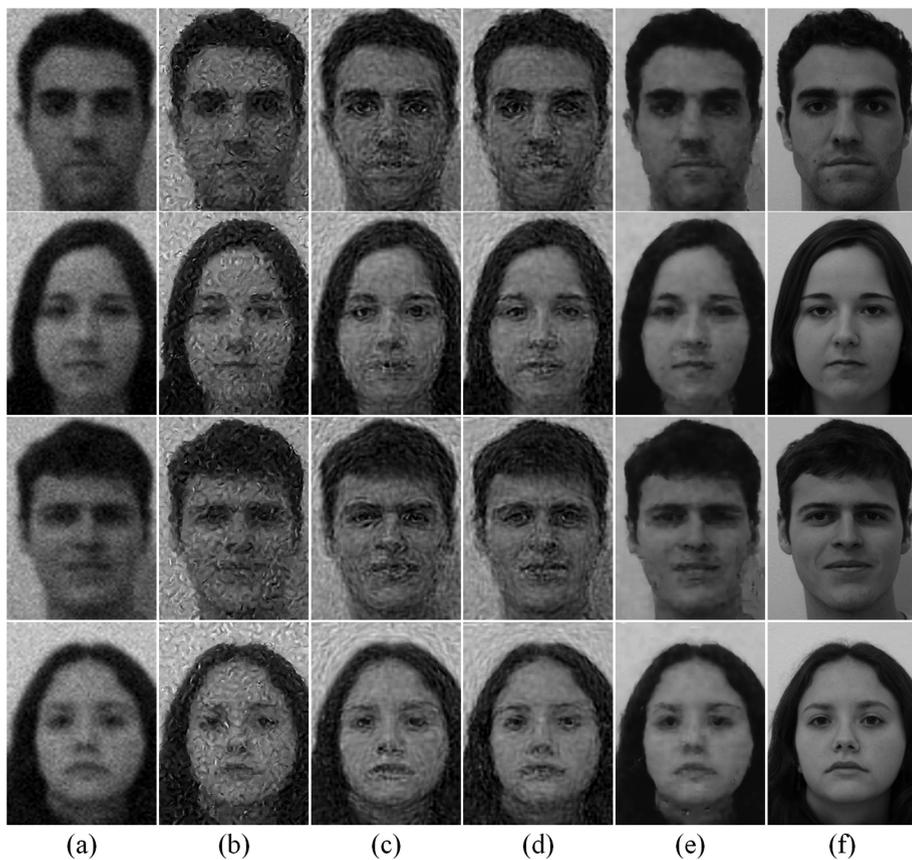


Fig. 5. Comparison of results based on different methods on FEI face database for noisy images ($\sigma = 10$). (a) Bicubic. (b) Chang's NE [7]. (c) Ma's LSR [8]. (d) Jiang's LcR [15]. (e) Proposed SSR. (f) Original HR faces.

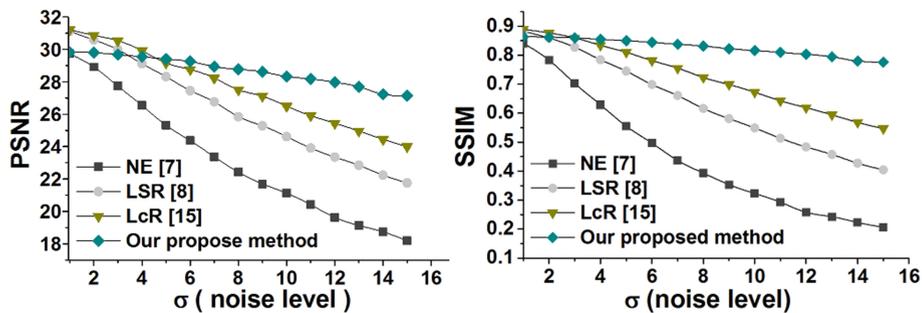


Fig. 6. Comparison of different methods: Average PSNR and SSIM with the noise level σ grows.

two curves are almost flat, which means the reconstruction quality is basically unchanged even the number of faces we use to reconstruct is reduced to half of the entire set. Therefore, we set M to 190 (half of the entire set) without affecting the quality in the following tests, while other methods still use the entire set with 380 training faces.

We conduct the following experiments under two unconformity degradation conditions: noise corrupted and smoothed by different kernels. NE [7], LSR [8] and LcR [15] methods are tested for better comparison.

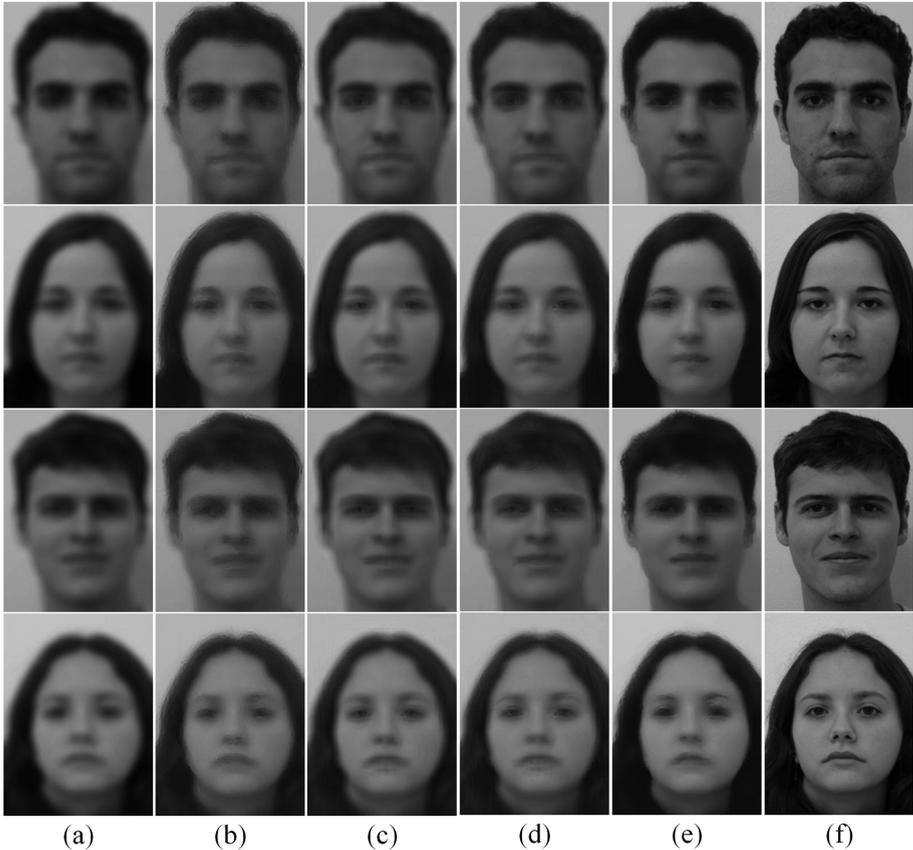


Fig. 7. Comparison of results with changed smooth kernel (Results best viewed with adequate zoom level where each face is shown with original size 360×260). (a) Bicubic. (b) Chang's NE [7]. (c) Ma's LSR [8]. (d) Jiang's LcR [15]. (e) Proposed SSR. (f) Original HR faces.

4.1 Robustness Against Noise

We add a zero-mean Gaussian noise ($\sigma = 1, 2, \dots, 15$) to test face to get the simulated noisy face. The smooth kernel is the same with the one in the training phase. Some randomly selected objects' results are shown in Figs. 4 and 5 when σ is set to 5 and 10.

The corresponding average PSNR and SSIM values are listed in Table 1. We can see that with the increase of noise level, the input face is no longer sparse. So the traditional SR methods can't distinguish the noise component from original images, the performance drop dramatically, and there are many artificial texture due to gaussian noise. But our SSR method is much less affected by noise and is able to restore a much clearer face. Relative to the second best LcR, PSNR gains of SSR approach 0.23 dB and 1.83 dB respectively for noise level 5 and 10. More results under different noise levels are shown in Fig. 6. As we can see, with the noise level continues to grow, SSR method will continue to widen the gap with the traditional methods.

Table 1. PSNR and SSIM comparison of different methods

Images	Metrics	NE [7]	LSR [8]	LcR [15]	Proposed SSR
$\sigma = 5$ Gaussian	PSNR	25.31	28.32	29.17	29.40
	SSIM	0.5549	0.7447	0.8102	0.8502
$\sigma = 10$ Gaussian	PSNR	21.13	24.61	26.50	28.33
	SSIM	0.3229	0.5489	0.6722	0.8162
Kernel changed	PSNR	27.86	28.93	29.30	28.45
	SSIM	0.8446	0.8570	0.8587	0.8414

4.2 Robustness Against Kernel Changes

We find the performance of SR methods decrease significantly even the kernel in reconstruction phase is slightly changed, compared to the one in training phase. However, this situation is the most common one in practical applications. Therefore, we decide to test different methods under the condition of inconformity kernels in training and reconstruction phase. As mentioned above, a fixed 20×20 gaussian lowpass filter with standard deviation of 4 is used in training phase. We change the standard deviation to 8 from 4 to make the input face more blurred than the faces in LR set.

According to Table 1, despite SSR method can't beat LSR and LcR in PSNR and SSIM values, results in Fig. 7 intuitively demonstrate that SSR method can create more details and sharper edges than others, which is more valuable in practical use. Performances of the first three methods are influenced a lot due to the different kernel, only the eye's region is artificially enhanced while other region is basically the same blurred as input face. But SSR can generate much clear facial edges like eyes, nose and facial contours, which proves its superiority in subjective visual quality.

5 Conclusion

In this paper, we have proposed a general parallel two-layer face hallucination framework to boost reconstruction speed and improve robustness against noise. Our method can exclude the unnecessary samples from over-complete training set using global similarity selection without quality loss. Then the local similarity representation

stage can make the method output satisfactory result under severe conditions. Experiments on FEI face database demonstrate this method can achieve better results under heavy noise conditions and gain good visual quality when the degradation process is changed. As a general framework, many other existing face hallucination schemes can also be incorporated into SSR method conveniently.

References

1. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: A comprehensive survey to face hallucination. *Int. J. Comput. Vis.* **106**, 9–30 (2014)
2. Freeman, W., Pasztor, E., Carmichael, O.: Learning low-level vision. *Int. J. Computer. Vis.* **40**(1), 25–47 (2000)
3. William, T.F., Thouis, R.J., Egon, C.P.: Example-based super-resolution. *IEEE Comput. Graph. Appl.* **22**(2), 56–65 (2002)
4. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(9), 1167–1183 (2002)
5. Dalley, G., Freeman, B., Marks, J.: Single-frame text super-resolution: a bayesian approach. In: *Proceedings of IEEE Conference on Image Processing*, vol. 5, pp. 3295–3298 (2004)
6. Yang, C., Liu, S., Yang, M.: Structured face hallucination. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1099–1106 (2013)
7. Chang, H., Yeung, D., Xiong, Y.: Super-resolution through neighbor embedding. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 275–282 (2004)
8. Ma, X., Zhang, J., Qi, C.: Position-based face hallucination method. In: *Proceedings of IEEE Conference on Multimedia and Expo*, pp. 290–293 (2009)
9. Ma, X., Zhang, J., Qi, C.: Hallucinating face by position-patch. *Pattern Recogn.* **43**(6), 3178–3194 (2010)
10. Matthew, A.T., Alex P.P.: Face recognition using eigenfaces. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–591 (1991)
11. Wright, J., Yang, A.Y., Ganesh, A., Shankar Sastry, S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. PAMI.* **31**(2), 210–227 (2009)
12. Yang, J., Tang, H., Ma, Y., Huang, T.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
13. Jung, C., Jiao, L., Liu, B., Gong, M.: Position-patch based face hallucination using convex optimization. *IEEE Signal Process. Lett.* **18**(6), 367–370 (2011)
14. Zhang, J., Zhao, C., Xiong, R., Ma, S., Zhao, D.: Image super-resolution via dual-dictionary learning and sparse representation. In: *Proceedings of IEEE International Symposium on Circuits Systems*, pp. 1688–1691 (2012)
15. Jiang, J., Hu, R., Han, Z., Lu, T., Huang, K.: Position-patch based face hallucination via locality-constrained representation. In: *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, pp. 212–217 (2012)
16. Jiang, J., Hu, R., Wang, Z., Han, Z.: Noise robust face hallucination via locality-constrained representation. *IEEE Trans. Multimedia* **16**, 1268–1281 (2014)
17. Wang, Z., Jiang, J., Xiong, Z., Hu, R., Shao, Z.: Face hallucination via weighted sparse representation. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2198–2201 (2013)

18. Wang, Z., Hu, R., Wang, S., Jiang, J.: Face hallucination via weighted adaptive sparse regularization. *IEEE Trans. Circ. Syst. Video Technol.* **24**, 802–813 (2013)
19. Weiss, Y., Freeman, W.T.: What makes a good model of natural images? In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)
20. Roth, S., Black, M.J.: Fields of experts. *Int. J. Comput. Vis.* **82**(2), 205–229 (2009)
21. Schmidt, U., Gao, Q., Roth, S.: A generative perspective on MRFs in low-level vision. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1751–1758 (2010)
22. Zhang, H., Zhang, Y., Li, H., Huang, T.S.: Generative bayesian image super resolution with natural image prior. *IEEE Trans. Image Process.* **21**(9), 4054–4067 (2012)
23. Liu, J., Ji, S., Ye, J.: SLEP: sparse learning with efficient projections (2010). <http://www.public.asu.edu/jye02/Software/SLEP>
24. FEI face database. <http://fei.edu.br/cet/facedatabase.html>