CONTEXT-PATCH BASED FACE HALLUCINATION VIA THRESHOLDING LOCALITY-CONSTRAINED REPRESENTATION AND REPRODUCING LEARNING

Junjun Jiang[1,2]*, Yi Yu*[2]*, Suhua Tang*[3]*, Jiayi Ma*[4]*, Guo-Jun Qi*[5]*, and Akiko Aizawa*[2]

[1]School of Computer Science, China University of Geosciences, Wuhan 430074, China
[2]Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan
[3]Department of Communication Engineering and Informatics, The University of Electro-Communications, Tokyo 182-8585, Japan
[4]Electronic Information School, Wuhan University, Wuhan 430072, China
[5]Department of Computer Science, University of Central Florida, Orlando, FL 32816, USA

ABSTRACT

Face hallucination, which refers to predicting a High-Resolution (HR) face image from an observed Low-Resolution (LR) one, is a challenging problem. Most state-of-the-arts employ local face structure prior to estimate the optimal representations for each patch by the training patches of the *same position*, and achieve good reconstruction performance. However, they do not take into account the *contextual information* of image patch, which is very useful for the expression of human face. Different from position-patch based methods, in this paper we leverage the contextual information and develop a robust and efficient *context-patch* face hallucination algorithm, called *Thresholding Locality-constrained Representation with Reproducing learning* (TLcR-RL). In TLcR-RL, we use a *thresholding* strategy to enhance the stability of patch representation and the reconstruction accuracy. Additionally, we develop a *reproducing learning* to iteratively enhance the estimated result by adding the estimated HR face to the training set. Experiments demonstrate that the performance of our proposed framework has a substantial increase when compared to state-of-the-arts, including recently proposed deep learning based method.

Index Terms— Face hallucination, super-resolution, context-patch, thresholding, reproducing learning

1. INTRODUCTION

Nowadays, surveillance cameras have been widely used in security and protection systems. They can provide very important clues about objects for solving a case, such as criminals. However, the object is so far away from the camera that the resolution of the interested face is too low to provide helpful information. Image super-resolution reconstruction is a class of techniques that can reconstruct a High-Resolution (HR) image with fine details from an observed Low-Resolution (LR) image, and thus transcending the limitations of an imaging system. Due to the highly underdetermined constraints and possible noise, super-resolution is a serious ill-posed problem and needs the prior information, such as piecewise smoothness, shape edges, local/nonlocal similar patterns, to regularize its solution. In this paper, we mainly focus on how to exploit the human face structure prior and super-resolve LR face image, also known as face hallucination.

Since the pioneer work by Baker and Kanade [1] and Liu *et al.* [2], face hallucination problem has increasingly been studied and attracted great research interests in recent decades [3]. Among them, position-patch based methods have attracted much interest. They assume that patches from the some position have similar contents and use all training image patches at the same position to represent the observation patch. For example, Ma *et al.* [4] is the first to use position constraint to regularize the reconstruction of patch and proposed a Least Square Representation (LSR) based method. To solve the unstable solution of LSR, Sparse Representation (SR) based methods have been developed by incorporating the sparsity constraint to the patch representation objective function [5]. In [6], Jiang *et al.* developed a Locality-constrained Representation (LcR) method which simultaneously incorporated sparsity as well as the locality into the patch representation task. Based on LcR method, they further improved the hallucination results by smooth technologies [7, 8] or by deep layer neighbor embedding [9].

However, aforementioned approaches neglect the contextual information of image patch, which has been proofed to be a very useful in the image denoising and classification tasks [10]. To incorporate the contextual information, the most simple and direct way is to enlarge the patch size as discussed in [11]. However, working with large patches rapidly

The research was supported by the National Natural Science Foundation of China under Grant 61501413 and 61503288, and JSPS KAKENHI Grant Number 16K16058.

978-1-5090-6067-2/17/$31.00

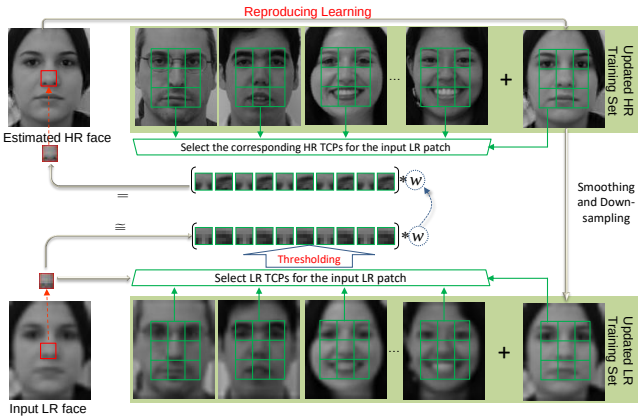


Fig. 1. Flow diagram of the proposed method. Given an LR patch (red patch), we first calculate its distance to the patches in LR Training Contextual Patches (TCPs), marked by green boxes. And then, we choose the similar contextual patches by hard thresholding to reconstruct the LR input patch and obtain the representation coefficient $\mathbf{w}$. Then, the target HR patch can be predicted by the linear combination of the corresponding HR TCPs with the same representation coefficient $\mathbf{w}$. By combing all the predicted HR patches, we can obtain the estimated HR face. In order to further improve the performance, we add the estimated HR face as well as its degenerated LR one to the training set, and repeat the *thresholding-representation-prediction* steps to update the target HR face iteratively, until the termination condition is reached.

becomes impractical since the training set size should grow exponentially with the patch size in order to guarantee appropriate patch reconstruction [10]. More recently, Dong *et al.* [12] developed SRCNN to learn a mapping from LR to HR space in an end-to-end manner. Kim *et al.* [13] further used very deep networks to do the super-resolution task, and demonstrated that deeper networks can achieve better performance because of large receptive field and high nonlinearities. Large receptive field means taking more contextual information (very large image regions) into account.

To address the aforementioned issues and utilize the contextual information without enlarging patch size, in this paper we propose to use all the patches in a large window around the observation patch and develop a novel face hallucination framework based on context-patch (Fig. 1). *It inherits the merits of working with small patches, while having the potential gains that large patches (large receptive field) bring.* Under this framework, we propose an effective and efficient patch representation based on Threshold LcR (TLcR), which can improve the stability of representation and the reconstruction accuracy. Additionally, we also develop an iterative enhancement strategy through Reproducing Learning (RL). These two improvements make the proposed TLcR-RL method robust to misalignment.

2. PRELIMINARIES

In learning based face hallucination methods, the training set is composed of HR and LR face image pairs. HR training set is denoted as $\mathscr{Y}_H = \{\mathbf{Y}_H^m\}_{m=1}^M$ and their LR counterpart is denoted as $\mathscr{Y}_L = \{\mathbf{Y}_L^m\}_{m=1}^M$, where $\mathbf{Y}_H^m$ and $\mathbf{Y}_L^m$ denote the HR and LR training images with index m, respectively, and M is the number of training sample. The primary task is to hallucinate an HR face image $\mathbf{X}_H$ from an observed LR face image $\mathbf{X}_L$ with the assistance of HR and LR training pairs, $\{\mathbf{Y}_H^m, \mathbf{Y}_L^m\}_{m=1}^M$.

In position-patch based methods, all HR and LR training images and the LR input image are divided into small overlapping patches, $\{\mathbf{y}_H^m(p)\}_p$, $\{\mathbf{y}_L^m(p)\}_p$ and $\{\mathbf{x}_L(p)\}_p$, respectively, p indicates the position information. For the LR test patch $\mathbf{x}_L(p)$ located at p, those methods tries to obtain the optimal representation coefficients $\mathbf{w}_L(p)$.

As a result, the target HR patch $\mathbf{x}_H(p)$ can be predicted by transforming the optimal representation coefficients obtained in the LR training space to the HR training space $\mathbf{x}_H(p) = \sum_{m=1}^{M} \mathbf{y}_H^m(p) w_L^m(p)$. Finally, based on their corresponding positions and simply averaging the overlapping regions, an estimation of the HR target face x_H can be obtained by concatenating all the hallucinated HR patches $\{\mathbf{x}_H(p)\}_p$. For notation simplicity, we remove the the term (p) for convenience in the following.

3. PROPOSED METHOD

3.1. Context-Patch based Face Hallucination

Different from the traditional position-patch based face hallucination methods, we leverage all the contextual patches around position p to construct the LR inout patch,

$$J(\mathbf{w}_L) = \left\| \mathbf{x}_L - \sum_{n=1}^{N} \mathbf{y}_{LTCP}^n w_L^n \right\|_2^2 + \tau\Omega(\mathbf{w}_L), \qquad (1)$$

where the first term is the data-fidelity, the second term is the prior about the reconstruction weights, the regularization parameter τ is used to balance their contributions, and N is the total number of the LR TCPs and can be calculated by cM. Here, c is the number of LR TCPs at one LR face image. In this paper, we define the contextual patches as all the patches within a window centered at position p. Thus, the number c can be obtain by the window size (ws), patch size (ps) and step size (ss), $c = \left(1 + \frac{ws-ps}{ss}\right)^2$. As for the step size, we fix it to 2, and the settings about the patch size, window size will be analyzed in the experimental section.

We denote by $\mathbf{Y}_{LTCP}$ the LR TCPs matrix, $\mathbf{Y}_{LTCP} = [\mathbf{y}_{LTCP}^1, \mathbf{y}_{LTCP}^2, ..., \mathbf{y}_{LTCP}^N]$, and $\mathbf{w}_L$ the corresponding $N \times 1$ dimensional representation coefficient vector, $\mathbf{w}_L = [w_L^1, w_L^2, ..., w_L^N]^T$, where $\mathbf{y}_{LTCP}^i$ is the i-th column of the

LR TCPs matrix $\mathbf{Y}_{LTCP}$ and w_L^i is its representation coefficient, $i = 1, \ldots, N$.

Based on the above definition, Eq. (1) can be rewritten as the following matrix form,

$$J(\mathbf{w}_L) = \|\mathbf{x}_L - \mathbf{Y}_{LTCP}\mathbf{w}_L\|_2^2 + \tau\Omega(\mathbf{w}_L). \tag{2}$$

Following the intuition that locality constrained linear coding encourages a local representation, and locality is more essential than sparsity, as locality must lead to sparsity but not necessary vice versa [14, 6], in this paper we choose the locality constraint to measure the priority of $\mathbf{w}_L$,

$$\begin{aligned} J(\mathbf{w}_L) &= \|\mathbf{x}_L - \mathbf{Y}_{LTCP}\mathbf{w}_L\|_2^2 + \tau \|\mathbf{d} \odot \mathbf{w}_L\|_2^2, \\ \text{s.t.} \quad & \sum_{i=1}^{N} w_L^i = 1, \end{aligned} \tag{3}$$

where "$\odot$" denotes the element-wise multiplication, $\mathbf{d}$ is a $N \times 1$ vector, $\mathbf{d} = [d_1, d_2, ..., d_N]^T$, with elements being the similarity between the LR input patch and all the LR TCPs. In this paper we simply use the Euclidean distance to measure the similarity, $d_n = \|\mathbf{x}_L - \mathbf{y}_{LTCP}^n\|_2, i = 1, \ldots, N$. The sum to one constraint follows the shift-invariant requirements of the locality constrained linear coding.

3.2. Thresholding Locality-constrained Representation

Through expanding the training size with contextual patches, the representation ability has been greatly improved. However, it is worth noting that when we set a large window to obtain many more contextual patches, the training sample size will increase to several times the size of the original one, i.e, $N \gg M$. As a result, the computational complexity of (3) will increase dramatically. In addition, using all LR TCPs will exacerbate the uncertainty of the sample space because many TCPs dissimilar to the input patch, thus increasing the difficulty of representing an input patch.

Inspired by the above observation, in this paper, we present an effective and efficient patch representation strategy by hard thresholding, which keep the coefficients of the K most similar training samples and these of other samples to zeros. By this hard thresholding strategy, it can simultaneously reduce the computational complexity and improve the reconstruction accuracy. Specifically, we add an extra constraint to the objective function (3) to select similar LR TCPs and eliminate the influence of dissimilar LR TCPs,

$$\begin{aligned} J(\mathbf{w}_L) &= \|\mathbf{x}_L - \mathbf{Y}_{LTCP}\mathbf{w}_L\|_2^2 + \tau \|\mathbf{d} \odot \mathbf{w}_L\|_2^2 \\ \text{s.t.} \quad & \sum_{i=1}^{N} w_L^i = 1 \ \text{ and } \ w_L^k = 0 \text{ if } k \notin \mathbb{C}_K(\mathbf{x}_L), \end{aligned} \tag{4}$$

where $\mathbb{C}_K(\mathbf{x}_L)$ denotes the indices of the K nearest neighbors of $\mathbf{x}_L$ in the LR TCPs. Through the added constraint, the value of w_L^i will be reduced to zero if $\mathbf{y}_{LTCP}^i$ is not in the K nearest neighbors of $\mathbf{x}_L$. In particular, we need only to find the K nearest neighbors of $\mathbf{x}_L$ in the LR TCPs, and use them to construct $\mathbf{x}_L$. Thus, Eq. (4) can be seen as a constrained least squares problem, which can be solved by an analytical solution [14, 6].

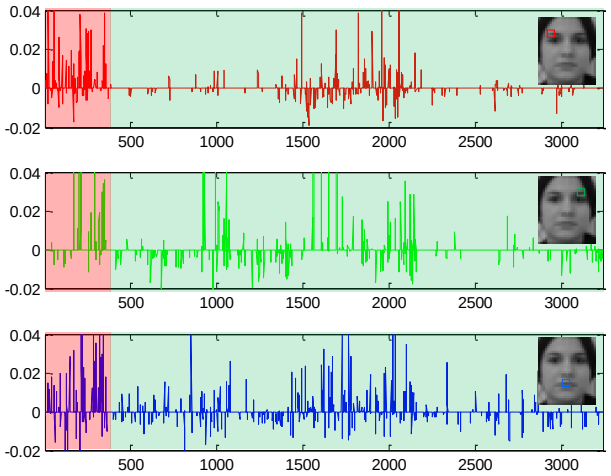


Fig. 2. Plots of the representation coefficients of three randomly selected LR input patches when using the K nearest neighbor LR TCPs. The horizontal axis is the index of LR TCPs, 1-360 are the indices of the position patches, while 361-3240 are the indices of surrounding contextual patches.

By incorporating the hard thresholding strategy to the context-patch based method, the proposed TLcR captures not only the information of position patches, but also that of the surrounding contextual patches. To demonstrate this point of TLcR, we design an experiment to see the contributions of position patches and surrounding contextual patches, *i.e.*, their representation coefficients. Particularly, we obtain the K dimensional ($K = 360$ in the experiment) coefficient vector corresponding to K nearest LR TCPs, and then rearrange them according to their indices in the LR TCPs matrix. As shown in Fig. 2, the first 360 values are from the position-patches, while the rest of values are form the TCPs except for the position-patches. In addition to the position patches, these surrounding contextual patches also have significant contributions to the test image patch, e.g., the representation coefficients of surrounding contextual patches (see the coefficients mark by the green region of each subfigure) are as large as these of position patches (see the coefficients mark by the red region of each subfigure). This indicates that the position-patch and the contextual patches simultaneously contribute to the representation of the LR text patch.

3.3. Reproducing Learning

For learning based methods, they have a weakness that requires a certain similarity between the training and the testing images. When the original HR face of the LR testing face is in the training set, the face hallucinating performance will be greatly enhanced. In contrast, when the input is very dissimilar to the training samples, the hallucination performance is

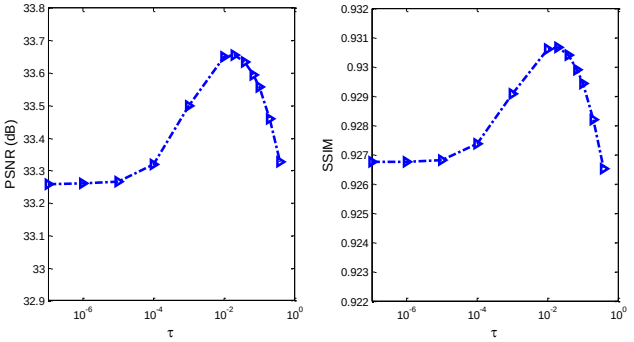


Fig. 3. The average PSNR and SSIM values of our proposed method according to the regularization parameter τ.

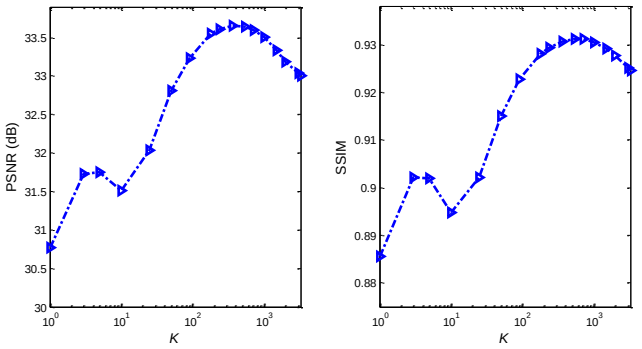


Fig. 4. The average PSNR and SSIM values of our proposed method according to the thresholding parameter K. The decrease around 9 is due to over-fitting on the LR patch with essentially 9-dimensional feature vector (3×3 pixels) [16].

poor. Inspired by this, in this paper we propose a reproducing learning method by adding the estimated HR image to the original training set. In this way, the "original" HR face (actually, the iteratively estimated HR face) of the LR test face is in the training set in the face reconstruction, and this will lead to 0.40 dB gain in term of PSNR (see the experimental section).

4. EXPERIMENT RESULTS

4.1. Experimental Settings

In the experiments, we conduct experiments on the public FEI Face Database [15] which has 400 face images. In the experiments, all the images are cropped to 120×100 pixels, and we choose 360 images as the training set, leaving 40 images for testing. Similar to [14, 6], the LR images are degenerated by an averaging smoothing filter of size 4×4 and 4× downsampling from corresponding HR images. For the sake of fairness, we set the patch size and overlap between neighbor patches of these local patch based methods to 12×12 pixels and 4 pixels, respectively.

4.2. Parameter Selection

There are mainly four parameters in the proposed method, *i.e.*, the regularization parameter τ, the thresholding parameter K, the window size, and the iteration value in reproducing learning. In the following, we will analyze the performance of our proposed method according to these four parameters.

In order to test the effectiveness of the locality constraint, we plot the PSNR and SSIM [17] values of our proposed method according to different regularization parameter τ (as shown in Fig. 3). When $\tau = 0$, the second term does not work, which can be regarded as the same case of neighbor embedding based method, the performance is not the best. With the increase of locality, much benefit can be gained. This implies that the locality constraint is very effective for patch representation. When the value of τ is between 0.01 and 0.1,

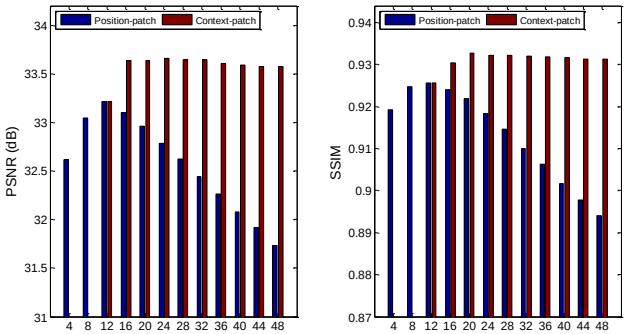


Fig. 5. The average PSNR and SSIM values of position-patch and context-patch based methods according to different patch sizes and window sizes, respectively.

the proposed method has obvious and stable improvements. In the following experiences, τ is set to 0.04.

Fig. 4 shows the objective values in terms of PSNR and SSIM according to different thresholding parameters K. It is worth noting that the total number of TCPs is 3240 (please refer to Section 1). When the thresholding parameter K is between 180 and 540, the proposed method can continually obtain a stable and good performance. Too small K could not reconstruct the testing patch well, while too large K will increase the difficulty of representing the testing patch, thus leading to uncertainty of representation. When K is 3240, PSNR and SSIM are 33.14 dB and 0.9259, which have 0.39 dB and 0.0045 decreases compared to the optimal setting, *i.e.* when K is 360. This implies the effectiveness of our proposed hard thresholding scheme.

To verify the effectiveness of context-patch based face hallucination scheme, we compare position-patch based and context-patch based methods. As shown in Fig. 5, the blue bars are the results of position-patch based method with various patch sizes, while the red bars are the results of context-patch based method with various window sizes. As for position-patch based method, it reaches the best results at 12×12 pixels and diminishes as the patch size becomes larg-

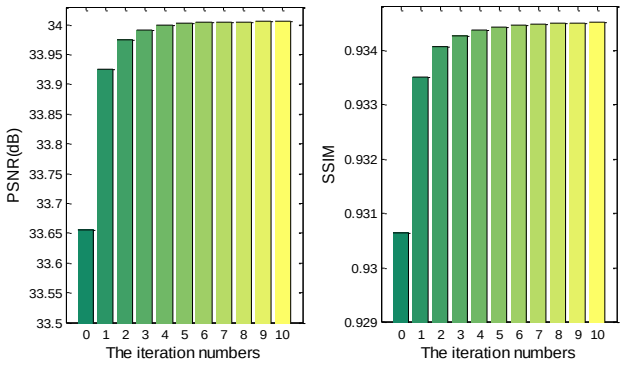


Fig. 6. Plots of the average PSNR and SSIM values with the increase of the iteration number in reproducing learning.

er. Note that when the patch size is 12×12 pixels and the window size is 12×12 pixels, the context-patch based method reduces to position-patch method. As for contextual-patch based method, by incorporating the contextual information (*e.g.*, set window size to 16), i.e., taking all the surrounding patches as the training patches, the performance of our proposed method has a significant improvement, *e.g.*, 0.42 dB and 0.0048 in terms of PSNR and SSIM, respectively. When the window size is larger than 16×16 pixels, there is a slight increase, and then it tends to decrease when window size is beyond 32. In all the following experiments, we set the window size to 16×16 pixels to balance the computational complexity and hallucination performance.

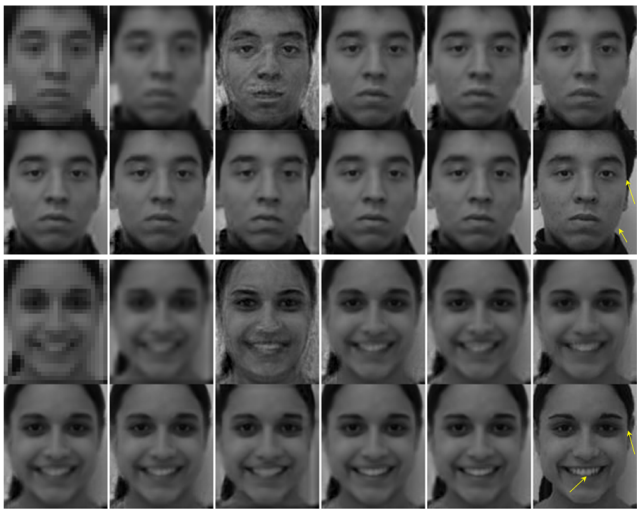


Fig. 7. Comparison of results based on different methods. From left to right, top to bottom: the input LR faces, results of Bicubic, Wang *et al.*'s method, NE [18], LSR [4], SR [5], LcR [6], LINE [9], SRCNN [12], TLcR, and TLcR-RL method. The last column is the original HR faces (Note that the effect is more obvious if the figure is zoomed.).

In order to mitigate the performance degradation of these learning based methods when the input LR face is not in the training set or is very dissimilar to the training samples, in this paper we propose the reproducing learning based enhancement scheme. By iteratively reproducing an updated training set, the proposed method can expect to achieve considerable improvements (Fig. 6). When the iteration number is 0, it reduces to the proposed TLcR method (without reproducing learning). By one time reproducing learning, it has an improvement of 0.30 in term of PSNR. With the increase of iteration, the gain is becoming more remarkable. The PSNR and SSIM improvements of TLcR-RL over TLcR are 0.39 dB and 0.0043. It is worth noting that the proposed method will reach a stable performance when the iteration number is larger than five and this also shows its quick convergence.

Table 1. Average PSNR (dB) and SSIM of different methods.

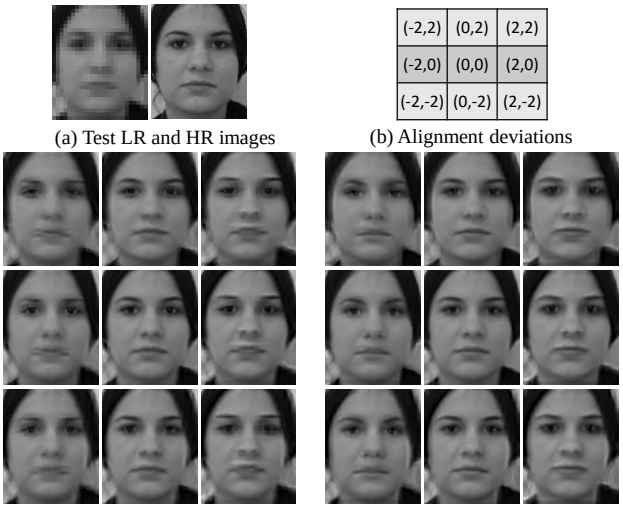
Methods	PSNR	SSIM
Bicubic	27.50	0.8426
Wang *et al.* [19]	27.57	0.7710
NE [18]	32.55	0.9104
LSR [4]	31.90	0.9032
SR [5]	32.11	0.9048
LcR [6]	32.76	0.9145
LINE [9]	32.98	0.9176
SRCNN [12]	33.13	0.9188
Our TLcR	**33.63**	**0.9304**
Our TLcR-RL	**34.03**	**0.9347**
Gain	0.90	0.0159

4.3. Comparison with state-of-the-arts

In order to verify the effectiveness of our proposed face hallucination framework, we compare the proposed method with several representative methods, including Wang *et al.*'s method [19], NE [18], LSR [4], SR [5], LcR [6], LINE [9], and SRCNN [12]. In addition, the results of the Bicubic interpolation method are given as baselines for comparison.

Objectively, we use PSNR and SSIM to evaluate the performance of different face hallucination methods. The PSNR and SSIM values of the reconstructed HR face images are shown in Table 1. It can be observed that the proposed method significantly improves the objective performance in terms of PSNR and SSIM, *e.g.*, 6.46 dB and 0.1637 better than Wang *et al.*'s global method [19], and 1.05 dB and 0.0171 better than the best position-patch based method, *i.e.*, LINE [9]. When compared with the recently proposed deep learning based SRCNN [12] (*i.e.*, the second best competing method in the experiments), the proposed method still has considerable gain, 0.40 dB and 0.0043 in terms of PSNR and SSIM, respectively.

Fig. 7 shows two groups of qualitative comparisons of our proposed method and other competing approaches. Due to the poor representation ability, Wang *et al.*'s method [19] has serious ghost effects and causes serious distortions. The HR



(a) Test LR and HR images (b) Alignment deviations

(c) Position-patch based method (d) Context-patch based method

Fig. 8. Hallucinated results of position-patch and context-patch based methods when the LR test face images have different levels of alignment deviations.

predictions of NE [18], LSR [4] and SR [5] are better than Wang *et al.*'s method [19], but have obvious artifacts around eyes and mouths. LcR [6], LINE [9] and recently proposed SRCNN [12] are very effective methods, which can maintain the primitive facial feature information. Through carefully observing the contours of the face, eyes, and noses, it can be seen that the hallucinated faces of our proposed method are much more visual pleasurable and similar to the ground truth.

4.4. Robustness to Misalignment

By introducing the contextual patches, the test patch will always find its true neighbors, *i.e.*, similar training patches, especially when the test face image is not aligned to the training samples. Obviously, for position-patch based method, it is not possible to select the true similar training patches when the test sample is not aligned to the training ones. To verify this point, we carry out one experiment when the test face image is not exactly aligned to the training samples. As shown in Fig. 8, (c) and (d) show the results of position-patch and context-patch based methods with different alignment deviations (b). When the test face image is well aligned (see the center faces of (c)), these two methods can well reconstruct the results. However, when there is alignment deviation, position-patch method has obvious ghost and artificial effects around the facial contours, eyes, and mouths. By contrast, the context-patch based method is much more visual pleasurable.

5. CONCLUSIONS

A novel framework that exploits contextual information of image patches for face hallucination is proposed. To improve the reconstruction accuracy, we develop a thresholding scheme to avoid being influenced by these dissimilar patches. Furthermore, we also develop an iterative enhancement strategy through reproducing learning. Experiments verify the effectiveness of our proposed locality constraint, thresholding scheme, contextual patch information, as well as the reproducing learning. We also show its robustness to misalignment. Comparison results with state-of-the-arts show that the effectiveness and superiority of the proposed algorithm.

6. REFERENCES

[1] S. Baker and T. Kanade, "Hallucinating faces," in *FG*, 2000, pp. 83–88.

[2] C. Liu, H. Shum, and C. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model," in *CVPR*, 2001, vol. 1, pp. 192–198.

[3] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *IJCV*, vol. 106, no. 1, pp. 9–30, 2014.

[4] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *PR*, vol. 43, no. 6, pp. 2224–2236, 2010.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE TIP*, vol. 19, no. 11, pp. 2861–2873, 2010.

[6] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE TMM*, vol. 16, no. 5, pp. 1268–1281, Aug 2014.

[7] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, and R. Hu, "SRLSP: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE TMM*, vol. 19, no. 1, pp. 27–40, Jan 2017.

[8] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *CYB*, vol. PP, no. 99, pp. 1–12, 2016.

[9] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE TIP*, vol. 23, no. 10, pp. 4220–4231, 2014.

[10] Y. Romano and M. Elad, "Con-patch: When a patch meets its context," *arXiv preprint arXiv:1603.06812*, 2016.

[11] L. Chen, R. Hu, Z. Han, Q. Li, and Z. Lu, "Face super resolution based on parent patch prior for vlq scenarios," *MTAP*, pp. 1–24, 2016.

[12] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE TPAMI*, vol. 38, no. 2, pp. 295–307, 2016.

[13] J. Kim, J K Lee, and K M Lee, "Accurate image super-resolution using very deep convolutional networks," *arXiv preprint arXiv:1511.04587*, 2015.

[14] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face hallucination via locality-constrained representation," in *ICME*, 2012, pp. 212–217.

[15] Carlos E. Thomaz and Gilson A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *IVC*, vol. 28, no. 6, pp. 902–913, 2010.

[16] M. Bevilacqua, A. Roumy, C. Guillemot, and ML. Alberi, "Low-complexity single-image super-resolution based on non-negative neighbor embedding," in *BMVC*, 2012, pp. 1–10.

[17] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE TIP*, vol. 13, no. 4, pp. 600–612, 2004.

[18] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *CVPR*, 2004, vol. 1, pp. 275–282.

[19] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE TSMC-C*, vol. 35, no. 3, pp. 425–434, 2005.