Context-Patch Face Hallucination Based on Thresholding Locality-constrained Representation and Reproducing Learning

Junjun Jiang, *Member, IEEE,* Yi Yu, Suhua Tang, *Member, IEEE,* Jiayi Ma, *Member, IEEE,* Akiko Aizawa, and Kiyoharu Aizawa, *Fellow, IEEE*

***Abstract*—Face hallucination is a technique that reconstruct high-resolution (HR) faces from low-resolution (LR) faces, by using the prior knowledge learned from HR/LR face pairs. Most state-of-the-arts leverage position-patch prior knowledge of human face to estimate the optimal representation coefficients for each image patch. However, they focus only the position information and usually ignore the context information of image patch. In addition, when they are confronted with misalignment or the Small Sample Size (SSS) problem, the hallucination performance is very poor. To this end, this study incorporates the contextual information of image patch and proposes a powerful and efficient context-patch based face hallucination approach, namely Thresholding Locality-constrained Representation and Reproducing learning (TLcR-RL). Under the context-patch based framework, we advance a thresholding based representation method to enhance the reconstruction accuracy and reduce the computational complexity. To further improve the performance of the proposed algorithm, we propose a promotion strategy called reproducing learning. By adding the estimated HR face to the training set, which can simulates the case that the HR version of the input LR face is present in the training set, thus iteratively enhancing the final hallucination result. Experiments demonstrate that the proposed TLcR-RL method achieves a substantial increase in the hallucinated results, both subjectively and objectively. Additionally, the proposed framework is more robust to face misalignment and the SSS problem, and its hallucinated HR face is still very good when the LR test face is from the real-world. The MATLAB source code is available at https://github.com/junjun-jiang/TLcR-RL.**

***Index Terms*—Image super-resolution, face hallucination, context-patch, position-patch, reproducing learning.**

I. Introduction

Face hallucination, which can be seen as a domain-specific super-resolution technology, is a technique to infer a High-Resolution (HR) face image, along with increasing the detailed face features, from low-resolution (LR) face images [1]. It has numerous applications for face recognition [2], [3], 3D face modeling [4], criminal detection [5], [6], and so on. From the pioneering work of [7], [8], many issues of face hallucination have been increasingly studied [9], [10]. Generally speaking, these methods all try to explore the implicit or explicit transformation between the LR and HR spaces with an additional training set with LR and HR face image pairs. Most methods in the literature fall into two main categories: Statistical model based global face methods and patch prior based local face methods. *A list of face hallucination resources collected by Jiang can be found at [11].*

Statistical model based global face methods leverage the face statistical models, such as PCA [12], locality preserving model [13], uniform space projection [14], [15] and Nonnegative Matrix Factorization (NMF) [16], to model the face image and execute face hallucination globally. They can well maintain the global structure of human face. However, their results lack detailed local face features and suffer from ghosting artifacts.

Considering that the human face structure is a significant prior, many face hallucination methods try to exploit the prior knowledge present in smaller patches [17], [18], [19], [20]. Among them, position-patch based methods have gained widespread attention in recent years. The common idea of these methods is to divide the global face into many small patches with predefined patch size and overlap, and use the training patches with the same position as the input one to construct the input patch. In this paper, our work is mainly concerned with this type of approach.

The Least Square Representation (LSR) method [21] is one of the representative position-patch based methods [17]. To address the problem that the solution of LSR is unstable, Sparse Representation (SR) based models have been developed by incorporating the sparsity regularization [16], [22], [23], [24]. However, SR methods overemphasize sparsity and neglect local similarity among the training samples, which is essential for exploiting the intrinsic non-linear manifold of the training sample space [25], [26]. The approach of [27] develops a Locality-constrained Representation (LcR) model which simultaneously adds the sparsity and locality constraints to the patch representation objective function, obtaining stable and reasonable representation coefficients. In order to alleviate the inconsistency of the LR and HR spaces, some works have been proposed to iteratively obtain the patch representation

The research was supported by the National Natural Science Foundation of China under Grants 61501413, 61503288 and 61773295, and was also partially supported by JSPS KAKENHI Grant Number 16K16058.

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and is also with the Peng Cheng Laboratory, Shenzhen, China. E-mail: jiangjunjun@hit.edu.cn.

Y. Yu and A. Aizawa are with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan ({yiyu, aizawa}@nii.ac.jp).

S. Tang is with the Department of Communication Engineering and Informatics, The University of Electro-Communications, Tokyo 182-8585, Japan (shtang@uec.ac.jp).

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (jyma2010@gmail.com).

K. Aizawa is the Department of Information and Communication Engineering, The University of Tokyo, Tokyo 113-8654, Japan (e-mail: aizawa@hal.t.u-tokyo.ac.jp).

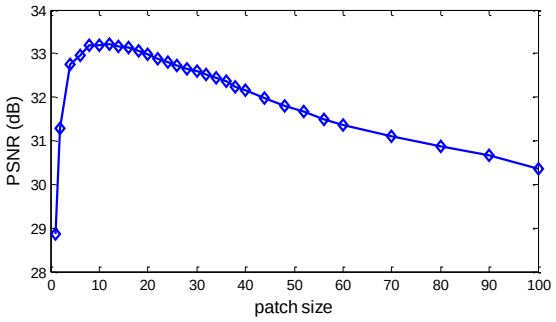


Fig. 1. Influence of the patch size over the position-patch based method [27]. Simply enlarging the patch size does not invariably bring performance improvement. It reaches the best performance when the patch size is set to 12 in the FEI database [36] with training sample size of 360.

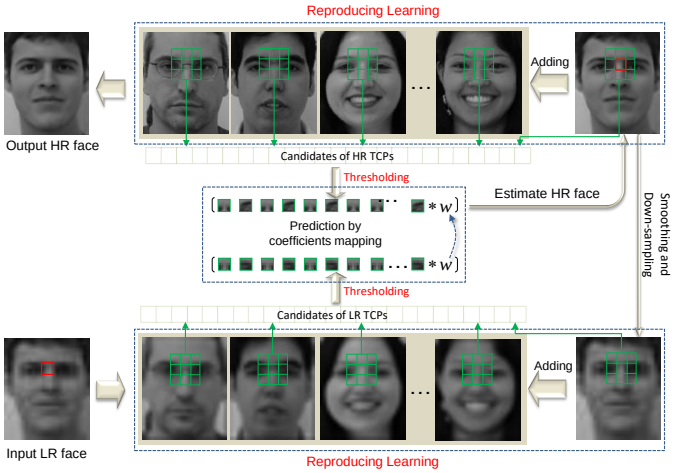


Fig. 2. Flow diagram of our proposed TLcR-RL based context-patch face hallucination framework. The face images marked with gray background are the HR and LR training sample pairs.

and perform neighbor embedding or learn the mapping in correlation spaces [15], [28], [29], [30]. Based on LcR, recently, the low-rank and self-similarity priors are also introduced to regularize patch representation in [31], [32], [33]. In [34], Pei *et al.* incorporated the gradient information of face image to further regularize the patch representation. In addition to face hallucination, the LcR algorithm has been also used to deal with pose and illumination problems in face hallucination and synthesis [9], [10], [35].

However, aforementioned local patch treatments mainly focus on the small patches and do not take into account the global nature, which has been verified to be beneficial to image description, image denoising and retrieval tasks [37], [38], [39]. To model the global nature of local patch based methods, the most direct way to incorporating the contextual information is to extend the patch as discussed in [40], [41]. The most extreme situation is to treat the entire face as a whole, using a global face-based approach. Another possible solution is to introduce a global reconstruction constraint in the image patch based method [42], [43], [30]. However, when the training sample size is fixed, it will become much more formidable to reconstruct a large patch or infer the global face image. In other words, because the training sample size should grow exponentially with the size of the image patch, it becomes impractical to present a too large image patch [39], [42]. This point is illustrated by Fig. 1 (to avoid the effect of overlap level under different patch sizes, we set the overlap pixel to the half of patch size). More recently, to reconstruct the latent HR image locally while thinking globally, DNNs, especially CNNs, have been applied to construct the mapping relationship between the LR images and their HR counterparts and shown strong learning capability and accurate prediction of HR images [44], [45]. For example, Dong *et al.* [44] developed a general image super-resolution method based on SRCNN. This is the very first attempt to use deep learning tools for image super-resolution reconstruction. The approach of Liu *et al.* [45] proposes to introduce the domain expertise to design a Sparse Coding based Network (SCN). Recently, R-DGN [46], CBN [47], LCGE [48], Attention-FH [49], FSRNet [50], and [51] are the most competitive approaches for face hallucination. They unitized very deep networks to model the relationship between the LR images and their HR correspondings, and verified that deeper networks can produce better results due to the large receptive field, which means considering more contextual information, *i.e.*, very large image regions.

A. Motivation and Contributions

To utilize the contextual information while without enlarging the patch size, this paper proposes to simultaneously consider all the patches in a large window centred at the observation patch that named *context-patch* and develop a context-patch based face hallucination framework. It inherits the merits of predicting with small local patches, while having the benefits of working with large patches (large receptive field). Based on Thresholding LcR (TLcR), the stability of representation and reconstruction accuracy can be improved. Observing that the reconstruction performance will be improved if there are some similar samples in the training set, we further advance an enhancement scheme via Reproducing Learning (RL), which puts reconstructed HR samples back to the training set and makes it easier to reconstruction the input image. As shown in Fig. 2, we illustrate the framework of our proposed context-patch based face hallucination algorithm. For a testing patch, which is marked by red box, on the input LR face image, we first extract the LR Context-Patches (TCPs), which are marked by green boxes, from the LR training set. Then, we calculate the distances between the input LR patch and TCPs, and choose the K nearest neighbor patches to reconstruct the input LR patch. Lastly, the output HR patch can be prediced by combining the corresponding HR TCPs with the representation coefficient $\mathbf{w}$ obtained in the LR training set. To promote the performance, we add back the hallucinated HR image to the training set, which can simulates the case that the HR version of the input LR face is present in the training set, and repeat the *thresholding-representation-prediction* steps to iteratively enhance the final hallucination result. In summary, the main contributions of this study are threefold.

- We introduce the concept of context-patch to expand the receptive field of the patch representation model. It not

only inherits the merits of predicting with small patches but also has the benefits of working with large patches. In addition, we combine the low-level pixel values and high-level position information to represent the image patch, thus further exploiting the contextual information.
- We develop a novel and robust image patch reconstruction method based on thresholding locality-constrained representation. It is inherited from the LcR method [27], but has the advantages of accurate patch representation and low computational complexity.
- We propose a face hallucination improvement strategy via reproducing learning. The estimated HR face image is iteratively reconstructed with a reproduced training set through adding the hallucinated HR image and its degenerated version to the training set. Experiments demonstrate its superiority some state-of-the-arts in term of both objective assessment and visual quality, especially when confronted with misalignment or the SSS problem.

The research reported in this paper is an extension of our preliminary work [52]. We highlight the significant differences between this research and [52] as follows: (i) to exploit much more contextual information of the patch images, we extent the pixel intensity based representation to the combination of low-level pixel values and high-level position prior (can be seen as the contextual information). (ii) the approach of [52] focuses only on controllable conditions. However, this research extends the application of TLcR-RL algorithm from controllable conditions to more intricate conditions, including both very limited training sample size (the SSS problem) and real-world image reconstruction. (iii) this research gives deep analysis on the motivations and advantages of introducing the contextual information, thresholding strategy, and reproducing learning, leading to a better understanding of why and how our method works.

B. Organization of This Paper

The rest parts of this study is organized as follows. In Section II, we give some notations and present the formulation of position-patch based methods. Section III presents the details of the TLcR-RL based face hallucination method followed by the improvement strategies of thresholding based patch representation and reproducing learning based iterative estimation. In Section IV, we report experimental evaluations of the context-patch based face hallucination framework and compare with some competitive algorithms. Finally, we conclude this paper and present the possible future work in the last section.

II. Preliminaries

In this paper, HR training set (consists of all HR training samples) is denoted as $\mathscr{Y}_H = \{\mathbf{Y}_H^m\}_{m=1}^M$ and their LR counterpart (consists of all LR training samples) is denoted as $\mathscr{Y}_L = \{\mathbf{Y}_L^m\}_{m=1}^M$, where $\mathbf{Y}_H^m$ ($\mathbf{Y}_L^m$) denotes the HR (LR) training samples with index m, and M is the size of training sample.

For position-patch methods, HR training images, LR training images, and the observed LR image are all divided into image patches according to the position information, $\{\mathbf{y}_H^m(p)\}_p$, $\{\mathbf{y}_L^m(p)\}_p$ and $\{\mathbf{x}_L(p)\}_p$, respectively, p is the position index. Given the LR testing patch $\mathbf{x}_L(p)$, the position-patch based method tries to utilize different constraints to regularize the representation coefficients $\mathbf{w}_L(p)$ in the LR training space:

$$J(\mathbf{w}_L(p)) = \left\| \mathbf{x}_L(p) - \sum_{m=1}^{M} \mathbf{y}_L^m(p) w_L^m(p) \right\|_2^2 + \tau\Omega(\mathbf{w}_L(p)), \tag{1}$$

where $\mathbf{w}(p)$ refers to the prior about the reconstruction weights. τ is a locality regularization parameter used to balance the contributions between the reconstruction errors and prior knowledge, *i.e.*, the closeness to the LR testing patch and the desired properties of the representation coefficients.

Based on the Locally Linear Embedding (LLE) [53] manifold learning assumption that HR and LR samples share similar local geometrical structure [54], which is characterized by the reconstruction coefficients in a neighborhood, these patch representation methods directly transform the LR representation coefficients to the HR space, *i.e.*, let $\mathbf{w}_H(p) = \mathbf{w}_L(p)$:

$$\mathbf{x}_H(p) = \sum_{m=1}^{M} \mathbf{y}_H^m(p) w_L^m(p). \tag{2}$$

Upon acquiring all the estimated HR face image patches $\{\mathbf{x}_H(p)\}_p$, the target HR face image can be calculated by placing all the estimated HR patches into original position and averaging each pixel from different reconstruction patches. For the simplicity of notation, we remove the term "(p)" in the following.

III. Context-Patch Based Face Hallucination

This section presents the proposed context-patch based face hallucination approach. First, we give the formulation of context-patch locality constrained representations. Then, we present the proposed thresholding approach to locality constrained presentation and reproducing learning strategy which aims to improve the reconstruction performance. Finally, we summarize the details of the face hallucination algorithm.

A. Context-Patch Representation

To construction the input LR face patch with the position of p, we use all the context-patches around position p to obtain its reconstruction coefficients through the following objective function:

$$J(\mathbf{w}_L) = \left\| \mathbf{x}_L - \sum_{n=1}^{N} \mathbf{y}_{LTCP}^n w_L^n \right\|_2^2 + \tau\Omega(\mathbf{w}_L), \tag{3}$$

where N is the number of LR TCPs. The value of N is determined by the window size (w), patch size(p) and step size (s):

$$N = M\left(1 + \frac{w-p}{s}\right)^2. \tag{4}$$

In this paper, we fix s to 2. The values of w and p are set experimentally.

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 representation vector,

$$\mathbf{w}_L = [w_L^1, w_L^2, ..., w_L^N]^T,$$

where $\mathbf{y}_{LTCP}^i$ is the LR TCP and w_L^i is its representation coefficient, $1 \leq i \leq N$.

With these definitions, we can rewrite the Eq. (3) as follows,

$$J(\mathbf{w}_L) = \|\mathbf{x}_L - \mathbf{Y}_{LTCP}\mathbf{w}_L\|_2^2 + \tau\Omega(\mathbf{w}_L). \tag{5}$$

Similar to [27], in this paper we employ the locality prior to regularize the representation of the input LR patch,

$$\begin{aligned} J(\mathbf{w}_L) &= \|\mathbf{x}_L - \mathbf{Y}_{LTCP}\mathbf{w}_L\|_2^2 + \tau \|\mathbf{d} \odot \mathbf{w}_L\|_2^2 \\ \text{s.t.} &\sum_{i=1}^{N} w_L^i = 1 \end{aligned} \tag{6}$$

where "$\odot$" denotes the element by element product of two vectors, $\mathbf{d}$ is a $N \times 1$ vector, $\mathbf{d} = [d_1, d_2, ..., d_N]^T$, with

$$d_n = \|\mathbf{x}_L - \mathbf{y}_{LTCP}^n\|_2, 1 \leq n \leq N. \tag{7}$$

In Eq. (6), the sum-to-one constraint $\sum_{n=1}^{N} w_L^n = 1$ is introduced to ensure that the reconstruction result is physically understandable.

By incorporating the locality-constraint, different LR TCPs will receive different penalties (or freedom). Specifically, those patches different from the LR testing one will be heavily penalized and have relatively small representation coefficient, while those patches similar to the LR testing one will receive relatively large representation coefficient, which is consistent with the intuitive understanding.

B. Thresholding Locality-constrained Representation (TLcR)

Compared with traditional position-patch based method, our context-patch based method can incorporate much more patches (N/M times) to construct the input patch. Thus, the representation ability of our method can be greatly promoted. However, this will also lead to two other problems: firstly, the multiplied increase in the training sample will lead to a rapid increase in computational complexity; secondly, many dissimilar image patches may be introduced to the training set, which will exacerbate the uncertainty. This is mainly because that when the number of training samples is far more than the dimension of LR patch, the patch representation problem is seriously ill-posed and they are many solutions. By selecting as few atoms (samples) as possible, it can expect to promote the patch representation stability as well as the reconstruction accuracy. The philosophy behind is same to the sparse representation and compressive sensing theory [55].

Based on the above observations, this paper proposes an effective and efficient image patch representation algorithm based on a thresholding strategy, which selects the K most similar training pathes to construct the input LR face image patch and sets the representation coefficients of other samples to zeros. Therefore, we can expect to greatly reduce the computational complexity and obviously improve the performance of the algorithm. In particular, we impose an additional

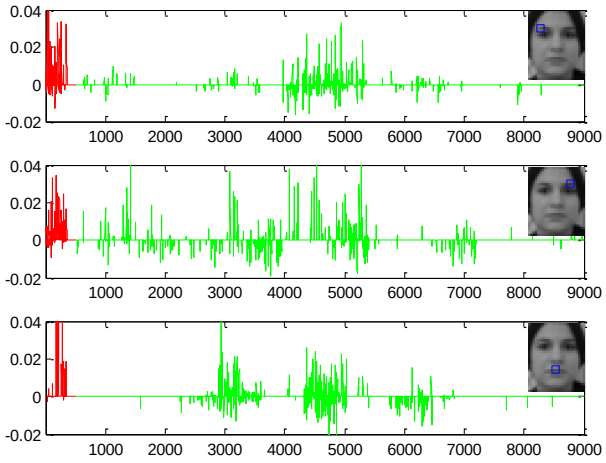


Fig. 3. Representation coefficients of some LR testing patches, which are marked by blue boxes. The red lines are the representation coefficients corresponding the position patches, while the green lines are representation coefficients corresponding the surrounding contextual patches. Here, the surrounding contextual patches denote all the TCPs except for the position-patches.

regularization term to Eq. (6) to select the K most similar LR TCPs pathes and discard the rest,

$$\begin{aligned} J(\mathbf{w}_L) &= \|\mathbf{x}_L - \mathbf{Y}_{LTCP}\mathbf{w}_L\|_2^2 + \tau \|\mathbf{d} \odot \mathbf{w}_L\|_2^2 \\ \text{s.t.} &\sum_{i=1}^{N} w_L^i = 1 \text{ and } w_L^k = 0 \text{ if } k \notin \mathbb{C}_K(\mathbf{x}_L), \end{aligned} \tag{8}$$

where $\mathbb{C}_K(\mathbf{x}_L)$ represents the indices of K nearest neighbor (KNN) of $\mathbf{x}_L$ in $\mathbf{Y}_{LTCP}$. By incorporating the addtional constraint, the coefficient is zero if $\mathbf{y}_{LTCP}^i$ is not in the set of KNN. Therefore, we directly use the KNN to construct $\mathbf{x}_L$,

$$\begin{aligned} J(\mathbf{w}_L^K) &= \left\|\mathbf{x}_L - \mathbf{Y}_{LTCP}^K\mathbf{w}_L^K\right\|_2^2 + \tau \left\|\mathbf{d}^K \odot \mathbf{w}_L^K\right\|_2^2 \\ \text{s.t.} &\sum_{k \in \mathbb{C}_K(\mathbf{x}_L)} w_L^k = 1, \end{aligned} \tag{9}$$

where $\mathbf{Y}_{LTCP}^K = \left\{\mathbf{y}_{LTCP}^k\right\}_{k \in \mathbb{C}_K(\mathbf{x}_L)}$ and $\mathbf{d}^K = \{d_k\}_{k \in \mathbb{C}_K(\mathbf{x}_L)}$ are the K nearest LR TCPs and the distances to these LR TCPs, respectively, $\mathbf{w}_L^K = \left\{w_L^k\right\}_{k \in \mathbb{C}_K(\mathbf{x}_L)}$ represents the representation coefficients of the K nearest LR TCPs. Eq. (9) is a convex quadratic problem and can be solved by an analytic optimal solution,

$$\mathbf{w}_L^K = (\mathbf{G}^T\mathbf{G} + \tau\mathbf{D}^2) \backslash ones(K, 1), \tag{10}$$

where $\mathbf{G} = \mathbf{x}_L \cdot ones(K, 1)^T - \mathbf{Y}_{LTCP}^K$, $\mathbf{D}$ is a $K \times K$ diagonal matrix with $\mathbf{D}_{kk} = d_k, k \in \mathbb{C}_K(\mathbf{x}_L)$, and $ones(K, 1)$ is a $K \times 1$ column vector whose elements are all ones. The final optimal representation coefficients are obtained by rescaling $\mathbf{w}_L^K$ to satisfy $\sum_{k \in \mathbb{C}_K(\mathbf{x}_L)} w_L^k = 1$.

Acquiring the optimal representation coefficients $\mathbf{w}_L^K$, the target HR patch y_L can be predicted by:

$$\mathbf{y}_L = \mathbf{Y}_{HTCP}^K\mathbf{w}_L^K. \tag{11}$$

where $\mathbf{Y}_{HTCP}^K = \left\{\mathbf{y}_{HTCP}^k\right\}_{k \in \mathbb{C}_K(\mathbf{x}_L)}$ denotes the corresponding K nearest HR TCPs.

Through finding KNN, it transforms the large linear system to a small one, reducing the computation complexity of the

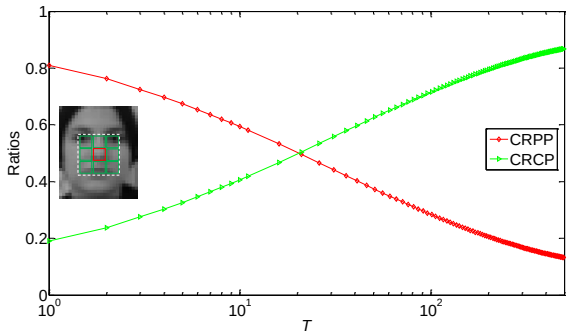


Fig. 4. Contribution analysis of position patches and surrounding contextual patches. $CRPP$: Contribution Ratio of Position-Patches, $CRCP$: Contribution Ratio of surrounding Contextual Patches. We plot the $CRPP$ and $CRCP$ according to different values of T.

linear system from $O(p^2KN + p^2K^3)$ to $O(p^2N^3)$, where $O(p^2KN)$ represent the additional complexity of $K - NN$ search in our method. We test the performance of with and without thresholding, we learn that the proposed thresholding scheme achieves 40 times faster than the original LcR method. In particular, the average running time of TLcR for one testing image is 13.8 seconds, while the LcR method will cost around 10 minutes.

To demonstrate the influence of position-patches and context-patches, in Fig. 3 we show where these patches that have non-zero coefficients come from. Clearly, we can see that these position-patches and surrounding contextual patches simultaneously contribute to the reconstruction of the testing patch. In addition, we also qualitatively test the contribution ratio of these two kinds of TCPs. Therefore, we define a metric called contribution ratio of position-patches ($CRPP$ for short):

$$CRPP(T) = \frac{\#\{ind_T \cap ind_{pp}\}}{T}, \tag{12}$$

where ind_T denotes the indices of the T most significant patches (the larger the values in the representation vector w are, the more significant the corresponding patches will be), ind_{pp} denotes the indices of the position patches, and $\#\{\cdot\}$ represents the cardinality of a set. Therefore, the contribution ratio of surrounding contextual patches ($CRCP$ for shot) is $CRCP(T) = 1 - CRPP(T)$. In Fig. 4, we plot the $CRPP$ and $CRCP$ according to T. We find that (i) when $T = 1$, $CRPP(1)$ is 81% and $CRCP(1)$ is 19%; (ii) when $T = 21$, $CRPP(21) \approx CRCP(21) \approx 50\%$; (iii) when $T = 360$, $CRPP(360)$ is 15% and $CRCP(360)$ is 85%; (iV) when $T = 500$, $CRPP(500)$ is 12% and $CRCP(500)$ is 88%. It demonstrates that in addition to the position patches, the contextual surrounding patches are also very significant for the patch representation, and not all significant patch necessarily come from the position-patches.

C. Reproducing Learning

The performance of learning based face hallucination methods (or general image super-resolution methods) usually depends on the distribution similarity between the training and testings samples, *i.e.*, the similarity between training and testing faces. If the HR face is preset in the training set, the reconstruction result is excellent. In contrast, when the original HR face of the input LR face is not in the training set and the input is dissimilar to samples in the training set, the performance of face hallucination algorithm will be very poor [56].

Inspired by the above observations, in this paper we propose to add the hallucinated HR face image to the training set, which can simulates the case that the HR version of the input LR face is present in the training set, and then perform TLcR based face hallucination with this newly generated training set (as show in Fig. 2). By reproducing learning, it will obtain 0.40 dB gain in term of Peak Signal-to-Noise Ratio (PSNR) and 0.0043 gain in term of Structural SIMilarity (SSIM) [57] (please refer to the experimental section).

In order to theoretically and technically explain the effectiveness of the proposed reproducing learning strategy, we give the following analysis. As a patch-based face hallucination approach, the hallucinated HR face is not necessarily the linear combination of training samples. By decomposing the entire face into smaller patches, the number of training patches will be greater than the patch size, the dictionary will become over-complete and it can reconstruct one patch with any content. Thats to say, in the absence of any other constraints, these patch-based methods can reconstruct any image which never appears in the database, e.g., a cat or a dog image, and thus introduce extra information to the training data. When the size of training set is smaller than the dimensionality of the face image, the face image to be reconstructed may not lay at the space spanned by the training samples. Therefore, if we put reconstructed HR samples (by a patch-based method) back to the training set, we can actually introduce some additional information in the sense that we can add an image that cannot be linear combination with the original training samples.

In addition, we also explain the effectiveness of the proposed reproduce learning strategy from the perspective of dictionary learning. As reported in the literature [58], [16], how to learn a representative dictionary is crucial in the image reconstruction and analysis problems. Sparse coding based dictionary learning method is the most successful dictionary learning technique [16], [59], [60], which aims at generating an over-complete dictionary with atoms that are linear combination with the training samples. Although the learned dictionary does not introduce additional information, it has a stronger reconstruction capabilities than the non-extended training set. As for our proposed method, putting the reconstructed HR samples back to the training set can be seen as generating a much more adaptive dictionary to the observed image. Thus, better reconstruction performance can be expected.

D. The Overall Algorithm

The complete face hallucination framework of our proposed TLcR-RL model is summarized as Algorithm 1. It should be noted that we extract the LR patch features by mean-removed pixel values. Moreover, in order to incorporate much more contextual information, we additionally incorporate the

Algorithm 1 Face Hallucination Based on TLcR-RL.

1: **Input**: HR and LR training face pairs, $\mathscr{Y}_H = \{\mathbf{Y}_H^m\}_{m=1}^M$ and $\mathscr{Y}_L = \{\mathbf{Y}_L^m\}_{m=1}^M$, and an LR input $\mathbf{X}_L$.
2: Divide the HR and LR training faces and the input LR face into small overlapping patches.
3: **for** each LR testing patch $\mathbf{x}_L$ **do**
4: Obtain the distance between $\mathbf{x}_L$ and all the N LR TCPs according to Eq. (7).
5: Select K most similar LR TCPs of $\mathbf{x}_L$.
6: Compute the optimal representation coefficients according to Eq. (10).
7: Construct the HR patch according to Eq. (11).
8: **end for**
9: Restore the HR face image by placing all predicted HR patches according to their positions and averaging each pixel from different reconstruction patches.
10: Reproduce a novel training set by adding the estimated HR image and its degenerated LR image to the original training set.
11: Repeat Step 2-Step 10 until the iteration number is reached.
12: **Output**: HR hallucinated face image $\mathbf{X}_H$.

position information, the vertical and horizontal coordinates of one patch, to the feature representation. This can be seen as the high-level information and has been successfully used in recovering the depth structure of human face [4]. Specifically, we leverage the relative coordinates to denote the position information.We use $\mathbf{x}_L = [\mathbf{x}_L; f \cdot p_x; f \cdot p_y]$ and $\mathbf{y}_L = [\mathbf{y}_L; f \cdot p_x; f \cdot p_y]$ to denote the feature representations of input LR patch and LR training patches. Here, f is the weight of the position information in the representation, and p_x and p_y are the vertical and horizontal coordinates of one patch. We experimentally set the value of f to 10 in our experiments, which will produce the best performance as shown in Fig. 5. For the HR images, we extract their high-frequency components, by subtracting the interpolated LR image, as the features. In the prediction stage, we add the estimated HR image into the Bicubic interpolated LR image. The aforementioned feature extraction and high-frequency prediction approach can improve the hallucinated results. In the experiments, we additionally found that when we use joint features of raw pixels and high-level patch position information (which can be seen as the contextual information of patches), e.g., simply combining them into a longer feature vector with a balancing scalar that controls the importance of the contextual information, the overall performance of our method will have about 0.20 dB improvement over the original low-level intensity based patch representation method.

IV. Experiments

This section evaluates the effectiveness of our proposed TLcR-RL method to face hallucination. Through these experiments, we can expect to answer following questions:

- Is the introduced contextual information helpful for face hallucination?

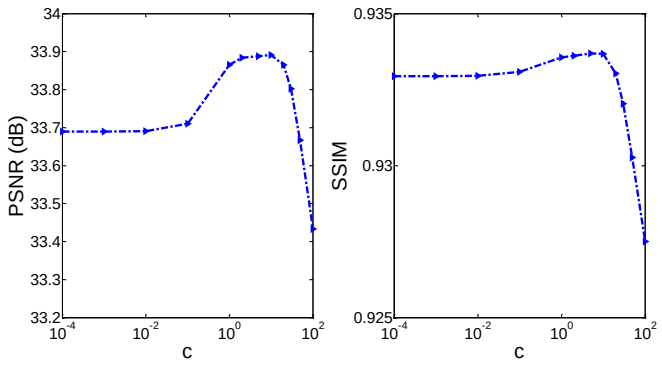


Fig. 5. Objective results according to different values of c. The performance will be stable when the value of c is between 1 and 20.

- How does TLcR-RL compare against state-of-the arts?
- How does the algorithm perform with different training sample sizes, and does it works well when confronted with the SSS problem?
- How robust is the algorithm to misalignment?
- Is the proposed method useful in real-world scenarios?

A. Experimental Settings

In the experiments, the public FEI database [36] is used. The HR faces are obtained by cropping the face image in FEI to 120×100 pixels. Similar to [27], 360 images are randomly selected as the training samples, while the rest 40 images are employed for algorithm testing. Thus, all the testing face images are absent in the training set. All face images are aligned in the FEI database. In practical, we can apply the automatic alignment methods and feature points matching methods [61], [62], [63] to preprocess face images. Similar to [27], [64], [31], we obtain the LR images by a filter (4×4 average smoothing) and 4× down-sampling. To balance the computational complexity and hallucination performance, the patch size and overlap pixels of all patch based methods are set to 12×12 pixels and 4 pixels, respectively, as in [27], [64], [31].

B. Model Analysis

In the proposed method, there are four parameters, *i.e.*, the balancing parameter τ and the thresholding number K in the objective function, the window size w, and the iterations in reproducing learning. By analyzing the above parameters on the performance of the algorithm, we can learn that the effectiveness of our proposed locality constraint, hard thresholding scheme, context-patch information, as well as the reproducing learning, are all testified.

1) Effectiveness of the locality constraint: In Fig. 6 we report the average PSNR and SSIM of TLcR-RL based face hallucination method according to τ. It can be seen that the performance of the algorithm rises first and then decreases. A too small or too large value of τ is not the optimal choice. When we set τ to [0.001, 0.1], the proposed TLcR-RL method will have stable improvements. Unless otherwise stated, τ is set to 0.04 in the all our experiences.

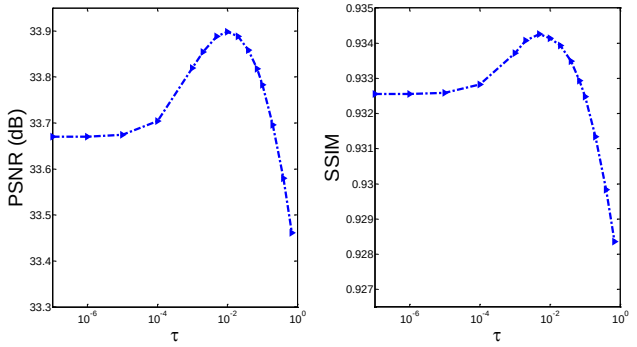


Fig. 6. Objective results in terms of average PSNR and SSIM of our proposed TLcR-RL based face hallucination method according to τ.

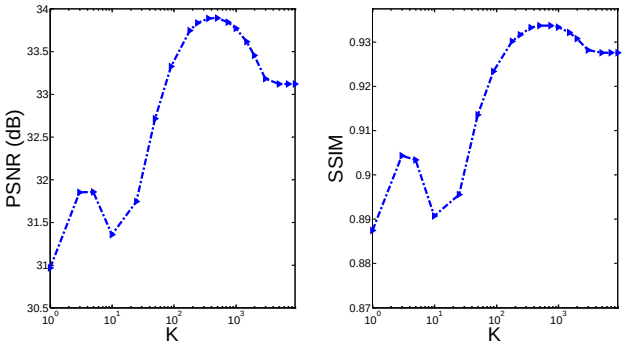


Fig. 7. Objective results of our proposed TLcR-RL based face hallucination method according to K. The decrease around 9 can be explained by the over-fitting on the input LR image patch [65].

2) Effectiveness of the hard thresholding scheme: Fig. 7 shows the average PSNR and SSIM according to K. Here, the total number of TCPs is 9000, which is calculated by substituting the patch size ($ps = 12$), window size ($ss = 20$), and step size ($ss = 2$) to Eq. (4), respectively. When the thresholding parameter K is between 180 and 1000, the proposed TLcR-RL method continually obtains a stable and good result. When the number of K is set too small, the limited training patches could not well reconstruct the input LR patch. In contrast, if the number of K is set too large, it will lead to the uncertainty of representation and increase the difficulty of representing a testing patch. Unless otherwise stated, K is set to 360 in the all our experiences.

3) Effectiveness of contextual patch information: When the window size is the same as the patch size, the proposed context-patch face hallucination method reduce to the position-patch based method. Therefore, to deomonstrate the effectiveness incorporating contextual information, we enlarge the window size to test the face hallucination performance. In Fig. 8, the blue and dark red bars show the objective performance of position-patch based and context-patch based methods according to the patch size and window size, respectively. When the patch size is with 12×12 pixels, position-patch based based method achieves the best performance. As we know, larger patch size can cover and convey more contextual information. However, when the training sample size is fixed, larger patch size does not mean better performance. This is mainly because the meaningful patch representation with large patch size calls for exponentially expanding the training set.

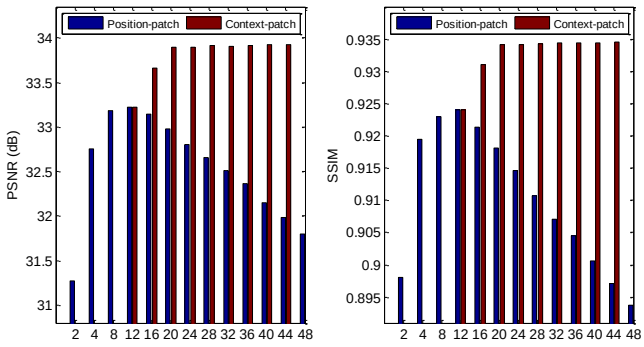


Fig. 8. Objective results of position-patch based (blue bars) and context-patch based (dark red bars) methods according to different patch and window sizes, respectively.

TABLE I
OBJECTIVE RESULTS IN TERMS OF AVERAGE PSNR (DB) AND SSIM OF THREE DIFFERENT PATCH REPRESENTATION STRATEGIES: ALL-PATCH (USING ALL THE PATCHES IN A FACE IMAGE), POSITION-PATCH, AND CONTEXT-PATCH.

Method	All-patch	Position-patch	Context-patch
PSNR	32.87	33.22	**33.86**
SSIM	0.9230	0.9256	**0.9336**
Gain	0.99	0.64	–
	0.0106	0.0080	–

By incorporating the contextual information, the performance of our proposed TLcR-RL based face hallucination method has a significant improvement, and reaches the stable performance at the window size of 20×20 pixels. When the window size is larger than 20×20 pixels, there is a very slight increase. This is reasonable because when the window size is too large, the extracted context-patches will be dissimilar to the input LR patch, and will be excluded by our proposed thresholding algorithm. To balance the computational complexity and face hallucination performance, we set the window size to 20×20 pixels in all our experiments.

Additionally, we consider an extreme situation when the window is set to the size of HR image, which means that we use the all patches on each face of the training set to represent the testing patch, the hallucination performance has a significant decrease when compared to the situation where the window size is 20×20 pixels or the position-patch method. Table I tabulates the objective results in terms of PSNR and SSIM of three different patch representation strategies, which shows again that the proposed context-patch based method is much more effective that position-patch based and all-patch based approaches. The portion-patch based method is the worst, and we attribute this to its unstable solution of patch representation when incorporating too much irrelevant training patches to the input patch. In other words, when we introduce all the image patches, the solution space of patch representation is too large (which further increasing the ill-posedness of the problem) and the locality constraint is not enough to find the optimal solution.

4) Effectiveness of reproducing learning: When the HR version of the testing LR face is absent or is dissimilar to samples in the training set, these learning based methods will work not very well to reconstruct the "out-of-samples".

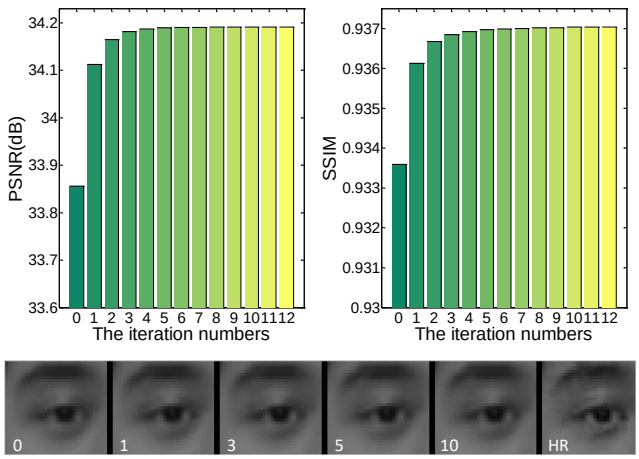


Fig. 9. Top: Plots of the objective results of the proposed TLcR-RL based face hallucination method with the increase of the iteration number in reproducing learning. Bottom: Hallucinated eye regions with different iterations and the original HR region for comparison.

TABLE II
THE AVERAGE PSNR (DB) AND SSIM PERFORMANCE OF LCR [27], TLCR AND TLCR-RL METHODS WITH DIFFERENT TRAINING SAMPLE SIZES (T.T.S.). AT THE LAST COLUMN OF EACH BLOCK, WE GIVE THE PERFORMANCE GAIN OF TLCR-RL OVER TLCR.

T.S.S.	PSNR				SSIM			
	LcR	TLcR	TLcR-RL	(Gain)	LcR	TLcR	TLcR-RL	(Gain)
360	32.76	33.86	34.19	**0.33**	0.9145	0.9336	0.9370	**0.0034**
300	32.67	33.78	34.15	**0.38**	0.9131	0.9326	0.9364	**0.0038**
200	32.44	33.58	33.97	**0.39**	0.9090	0.9305	0.9347	**0.0042**
100	31.75	33.08	33.53	**0.45**	0.8982	0.9253	0.9305	**0.0052**
75	31.42	32.91	33.38	**0.47**	0.8922	0.9238	0.9291	**0.0053**
50	30.61	32.54	33.12	**0.58**	0.8763	0.9199	0.9265	**0.0066**
20	28.00	31.54	32.40	**0.86**	0.8186	0.9085	0.9189	**0.0104**
10	25.66	30.69	31.87	**1.18**	0.7492	0.8977	0.9117	**0.0140**
5	22.07	29.95	31.70	**1.75**	0.6228	0.8937	0.9082	**0.0145**

We propose a novel improvement strategy named reproduce learning. It iteratively perform the face hallucination and training set emendation (by adding the estimated HR face to the training set). Fig. 9 shows the performance according to the iteration numbers. When the iteration number is 0, it means that no estimated HR face is added to the training set. By one time reproducing learning, it has a performance improvement of 0.25 dB in term of PSNR. As the number of iterations increases, the gain becomes more and more significant. At the bottom of Fig. 9, we show the hallucinated eye regions of one LR test face with different iterations. As shown, TLcR gives the smoother result, while TLcR with reproducing learning can recover most of the detailed feature. With the increase of the iterations, the hallucinated HR eye images are much more similar to the original eye region. It reaches a stable performance after a few iterations, *e.g.*, five time, which indicates the proposed method has a quick convergence.

To further testify the effectiveness of the reproducing learning, we give the performance gains of TLcR-RL over TLcR when they are confronted with the SSS problem, in which case it is more likely that no similar samples can be found in the training set. In Table II, we give the PSNR and SSIM performance of TLcR-RL and TLcR with different training sample sizes. Note that the results of LcR method[27] are also given as baselines for better comparison. With the decrease of the training sample size, the performance gains of TLcR-RL over TLcR become more and more obvious. In the very

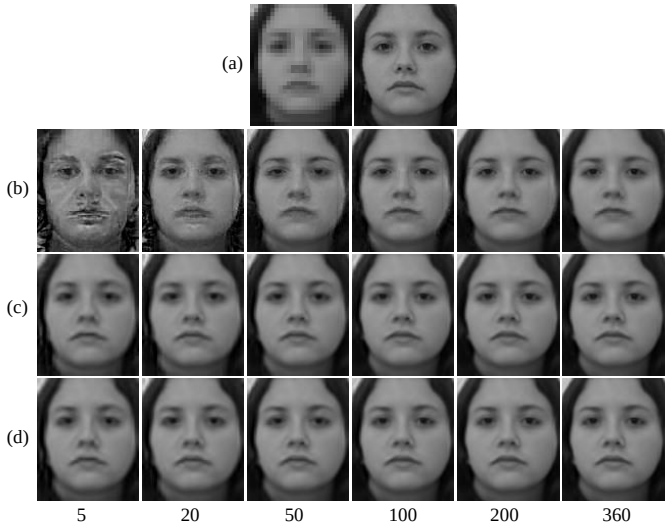
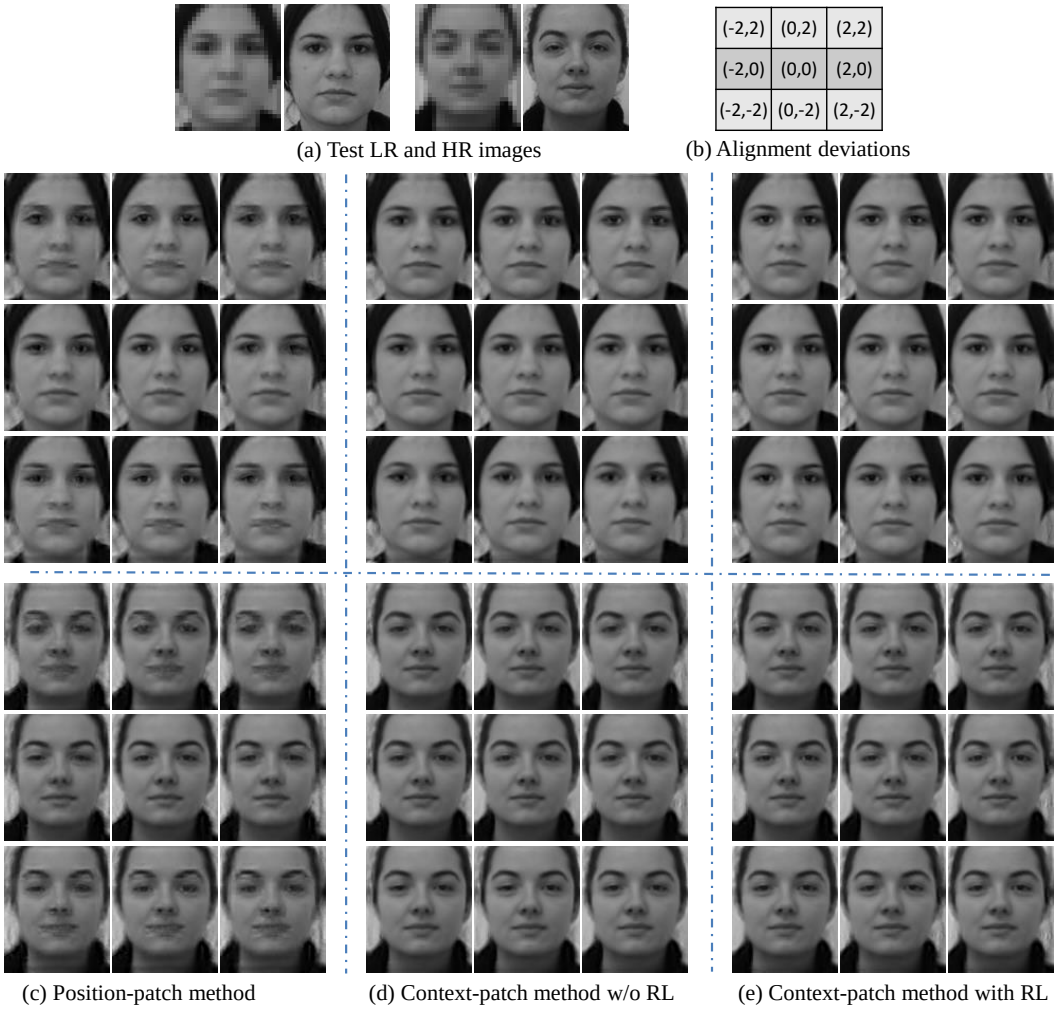


Fig. 10. Visual face hallucination results of LcR, TLcR and TLcR-RL with different training sample sizes. (a) is the input LR face and ground truth. (b), (c) and (d) are the results of LcR, TLcR and TLcR-RL methods, respectively. The numbers under the last row indicate the training sample sizes.

Fig. 11. Visual comparisons of two groups of hallucinated results by different face hallucination methods. For each group (every three rows), from left to right and top to bottom: the LR testing faces, hallucinated results of Bicubic interpolation method, Wang *et al.*'s method [12], NE [54], LSR [17], SR [16], LcR [27], LINE [64], DRP [43], LCDLRR [31], SCN [66], SRCNN [44], TLcR, and TLcR-RL. The last column is the HR ground truth.



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

(c) Position-patch method (d) Context-patch method w/o RL (e) Context-patch method with RL

Fig. 12. Two groups of visual face hallucination results of position-patch and context-patch based methods with misalignment LR input. (a) are two group of test LR and HR images. (b) denotes the alignment deviations of the test LR face in pixel. (c) is the hallucinated results by position-patch method, while (d) and (e) are the results of and context-patch based methods without reproducing learning and with reproducing learning, respectively.

extreme situation, *i.e.*, there are only five training samples in the training set, the performance gains of TLcR-RL over TLcR reach 1.75 dB (in term of PSNR) and 0.0145 (in term of SSIM). When compared with LcR [27], which is the most competitive position-patch based method, TLcR-RL is 9.63 dB better than LcR [27]. Fig. 10 shows the reconstructed HR images under several typical training sample sizes. Even in the case of only five training samples, the proposed TLcR-RL still performs very well, and its hallucinated face is much clearer than that of TLcR. In contrast, when the training sample size is less than 100, it is difficult for LcR [27] to reconstruct a pleasant result.

C. Comparison Results

In this section, we compare the proposed TLcR-RL with some state-of-the-arts, including Wang *et al.*'s global face method [12], NE [54], LSR [17], SR [16], LcR [27], LINE [64], LCDLRR [31], Shi et al.'s Dual Regularization Priors (DRP) based method [43], SCN [66] and SRCNN [44]. In addition, results of the Bicubic interpolation are the baselines. Note that NE [54], SCN [66], and SRCNN [44] are proposed for general image reconstruction instead of face reconstruction. In the experiments, we evaluate the NE [54] and SRCNN [44] models by the 360 HR and LR training pairs described in Section IV-B. As for SCN [66], we use the trained models by the authors directly to test its performance on face images. We carefully tune the parameter settings for other competitive methods to obtain their optimal performances. As for [12], the variance accumulation contribution rate is set to 99%. In NE [54], neighbor number is set to 75. We set error tolerance to 1.0 for SR [16]. In LSR [17] and LcR [27], the regularization parameters are set to 1e-6 and 0.04, respectively. In LINE [64], the neighbor number, the locality regularization parameter, and the iteration number are set to 100, 1e-4, and 5, respectively. As for SRCNN [44], we use the same image degradation as in previous methods, and the parameters are learned by the deep model. The results of LCDLRR [31] and DRP [43] are provided by the corresponding authors.

In Table III, we give the PSNR and SSIM performance of different face hallucination methods. We observe that TLcR-RL improves the objective performance *e.g.*, 1.05 dB and 0.0164 better (in terms of PSNR and SSIM) than the second best method, *i.e.*, LCDLRR [31]. We also compare with two deep learning methods, SCN [66] and SRCNN [44], the per-

TABLE III
THE OBJECTIVE RESULTS IN TERMS OF AVERAGE PSNR (DB) AND SSIM OF DIFFERENT METHODS. THE BEST AND SECOND BEST RESULTS ARE MARKED IN RED AND BLUE, RESPECTIVELY.

Methods	PSNR	SSIM
Bicubic	27.50	0.8426
Wang *et al.* [12]	27.57	0.7710
NE [54]	32.55	0.9104
LSR [17]	31.90	0.9032
SR [16]	32.11	0.9048
LcR [27]	32.76	0.9145
LINE [64]	32.98	0.9176
LCDLRR [31]	33.14	0.9206
DRP [43]	32.84	0.9292
SCN [66]	32.05	0.9048
SRCNN [44]	33.13	0.9188
Our TLcR	33.86	0.9336
Our TLcR-RL	34.19	0.9370
(Gain)	**1.05**	**0.0078**

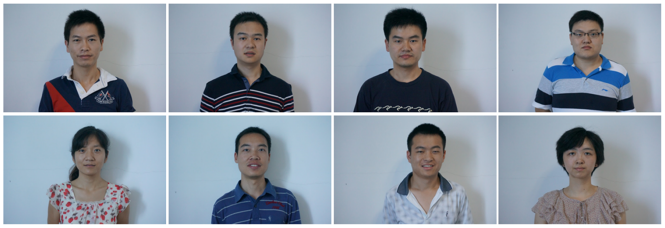


Fig. 13. Eight real-world images used to test the performance of different face hallucination methods. These images captured by an HD camera in the normal night condition. The eight face images are indexed as Img1 to Img8 in the following.

formance gain of our method over these two methods are still considerable. It should be noted that SRCNN [44] is retrained on the FEI database, so it can obtain better performance than SCN [66] that used the model trained by general images.

Fig. 11 shows qualitative comparisons of TLcR-RL and other approaches on four testing images. From the visual results, we see that PCA based global face method [12] has serious ghosting effects, and its results are dissimilar to the ground truth. The HR predictions of NE [54], LSR [17] and SR [16] are better than Wang *et al.*'s method [12], but have obvious artifacts around eyes and mouths. LcR [27], LINE [64], LCDLRR [31], DPR [43], and recently proposed deep learning methods, SCN [66] and SRCNN [44], are excellent methods, which can produce reasonable results that are similar to the ground truth. By carefully observing the contours of the face, eyes and nose, it can be seen that the resultant HR face images of TLcR-RL are more enjoyable and more similar to the original HR face images. In summary, the proposed TLcR-RL method demonstrates powerful hallucination ability quantitatively and qualitatively.

D. Robustness to Misalignment

To explore the contextual information, the context-patch based method has to introduce diverse training face image. Therefore, it will become very difficult for these position-patch based methods to find the true neighbors, *i.e.*, similar training patches, for the input testing patch, especially in condition that the input face is not well aligned to the training samples. In order to demonstrate this, we conduct one subjective experiment when the observed face image is not well aligned to faces in the training set. In Fig. 12, (a) is two well aligned faces, (b) is the alignment deviations of the test LR face in pixel, (c) shows the results of position-patch based method with different alignment deviations (b). Fig. 12(d) and (e) are the results of and context-patch based methods without reproducing learning and with reproducing learning, respectively. When the input LR face is well aligned to the training samples, both methods can well construct the target HR images. However, when the test face image has different alignment deviations (see Fig.12 (b), (0,0) indicates the observed face is well aligned to the faces in the training set), the reconstructed HR face of position-patch method has obvious ghosting effects. In contrast, the proposed context-patch based method can produce clear and shape edges. When compared with the hallucinated results with and without reproducing learning, we can see that the latter can well capture the facial details (please refer to the eye regions of these two methods). This once again proves the validity of the proposed reproducing learning algorithm.

E. Hallucinating with Real-World Images

In this subsection, we conduct one another experiment to demonstrate the effectiveness and the advancement of the proposed algorithm with some real-world face images that are very different from the face image in FEI database. We capture eight high-definition images as shown in Fig. 13.

Firstly, the commonly used automatic face detection algorithm [67] is applied to detect the face regions in the captured HR images, and then we align the detected faces to the mean face by the detected two points of eye centers[1]. Then, they are cropped to 120×100 pixels, which are the ground truth HR faces (see the last column of Fig. 14). In our experiments, the LR test faces are obtained by the same way as in Section IV-B (see the first column of Fig. 14). Here we only compare with two most representative local position-patch based methods, *e.g.*, LcR [27] and LINE [27], one global reconstruction constraint patch-based method, DRP [43], and the deep learning based method, *e.g.*, SRCNN [44], for their representative and good performance. The middle seven columns are the hallucinated results by Bicubic, LcR [27], LINE [64], SRCNN [44], TLcR and TLcR-RL. Note that, for these color face images, we change the input face images from RGB space to YUV space firstly, and then reconstruct them in the luminance component. This is mainly because humans are more sensitive to illuminance changes. The hallucinated eyes and face contours by SRCNN [44] are blurry and dirty. LcR [27] and LINE [64] produce some ghosting effects around the eyes, mouths and face contours. DRP [43] and the proposed can well maintain the face contours. Fig. 15 plots the results of these methods on the eight testing images. The proposed methods also show the best objective results. By further examination, it can be seen that the improvement on SSIM is much more obvious than that on PSNR, which indicates that our face hallucination model pays more attention to the face structure information and the hallucinated results are

[1]The automatic face detection algorithm outputs the coordinates of eye corner. We then use the mean coordinates to represent the eye center.

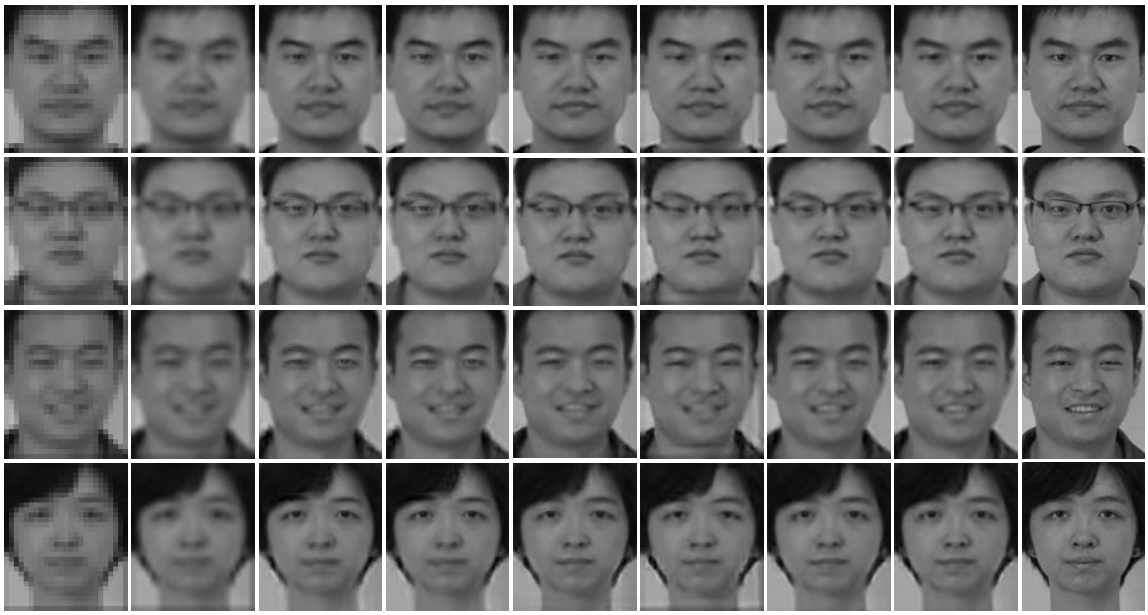


Fig. 14. Visual face hallucination results of the four real-world images: Img3, Img4, Img7, and Img8. From the first column to the last column: the LR test faces, results of Bicubic, LcR [27], LINE [64], DRP [43], SRCNN [44], TLcR and TLcR-RL, the ground truth HR faces.

much more consistent with visual perception. From the PSNR results of Img3 and Img4, we find that TLcR is worse than SRCNN [44]. However, the hallucinated results (see the first two rows of Fig. 14) of TLcR is much better than that of SRCNN [44]. This observation is consistent with the SSIM results. The SSIM results of DRP [43] are very competitive and this demonstrates its ability in maintaining the face structures. The above experiments indicate the effectiveness of the proposed TLcR-RL method in the real-world condition.

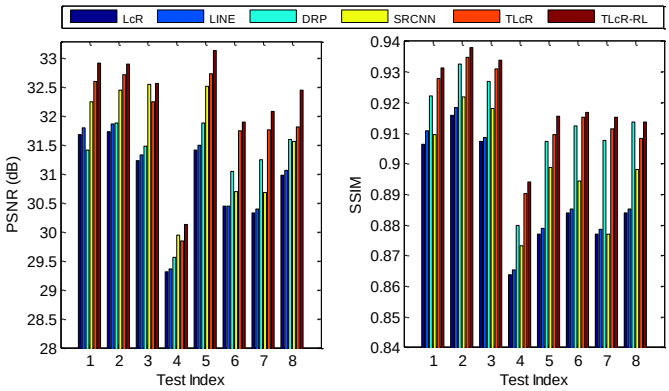


Fig. 15. PSNR and SSIM results of five competitive methods on the eight testing images. The average PSNR of these methods are 30.90 dB, 30.98 dB, 31.58 dB, 31.94 dB, and 32.26 dB, while the average SSIM of these methods are 0.8895, 0.8915, 0.8989, 0.9160, and 0.9198.

V. Main Findings and Future Work

In this paper we introduce a context-patch based face hallucination approach that can fully exploit contextual information of image patches. Different from conventional position-patch based approaches, which use only the training patches from the same position as the testing patch for reconstruction, the proposed method leverages the contextual information and uses the TCPs to obtain a better representation. In order to improve the accuracy of reconstruction, we have developed a hard threshold scheme to avoid being affected by these dissimilar training patches. In addition, we have also developed an iterative enhancement strategy to improve the estimation results by reproducing learning. The experiment verifies its robustness to misalignment and the SSS problem. Comparison results with some competitive approaches, including two deep learning based super-resolution methods, show that the hallucinated face images of our proposed approach have finer and more detailed features over state-of-the-arts.

In TLcR-RL, the representation coefficients of the input sample can be seen as the filter responses by a set of filters (or basis, *i.e.*, the training samples in our method), while the thresholding can be seen as the output of an activation function. The overlapped patch averaging strategy can be regarded as the filtering on a set of feature maps by some predefined filters. The above three operations can be formed as a convolutional layer. By incorporating reproducing learning, the proposed TLcR-RL is very similar to the form of DNNs. Therefore, how to combine the structure prior and the very efficient deep learning algorithms will be our first concern in the future.

In our experiments, we mainly focus on the reconstruction with frontal portrait in well controlled conditions, how to extend the proposed model to uncontrolled conditions, such as variety in poses, expression and illumination, will be our second future work.

Acknowledgment

We would like to thank Dr. Gao, the first author of [31], for his kind providing of the results of LCDLRR algorithm. We would like to thank the authors of [17], [16], [66], [44], and DRP [43] for their kind sharing of their source codes.

REFERENCES

[1] 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.

[2] W. Zou and P. Yuen, "Very low resolution face recognition problem," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 327–340, Jan 2012.

[3] J. Shi and C. Qi, "From local geometry to global structure: Learning latent subspace for low-resolution face image recognition," *IEEE Signal Proc. Let.*, vol. 22, no. 5, pp. 554–558, May 2015.

[4] S. Yang, J. Liu, Y. Fang, and Z. Guo, "Joint-feature guided depth map super-resolution with face priors," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 399–411, 2018.

[5] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, 2017.

[6] F. C. Lin, C. B. Fookes, V. Chandran, and S. Sridharan, "Investigation into optical flow super-resolution for surveillance applications," 2005.

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

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

[9] X. Ma, H. Song, and X. Qian, "Robust framework of single-frame face superresolution across head pose, facial expression, and illumination variations," *IEEE Trans. Hum.-Mach. Syst.*, vol. 45, no. 2, pp. 238–250, 2015.

[10] L. Liu, S. Li, and C. L. P. Chen, "Quaternion locality-constrained coding for color face hallucination," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1474–1485, 2018.

[11] J. Jiang, "Face hallucination benchmark," https://github.com/junjun-jiang/Face-Hallucination-Benchmark, accessed Aug 19, 2018.

[12] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. Syst. Man Cybern. Part C-Appl. Rev.*, vol. 35, no. 3, pp. 425–434, 2005.

[13] S. W. Park and M. Savvides, "Breaking the limitation of manifold analysis for super-resolution of facial images," in *ICASSP*, vol. 1, Apr 2007, pp. I–573–I–576.

[14] H. Huang, H. He, X. Fan, and J. Zhang, "Super-resolution of human face image using canonical correlation analysis," *Pattern Recogn.*, vol. 43, no. 7, pp. 2532–2543, 2010.

[15] L. An and B. Bhanu, "Face image super-resolution using 2D CCA," *Signal Proc*, vol. 103, pp. 184–194, 2014.

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

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

[18] Y. Li, C. Cai, G. Qiu, and K.-M. Lam, "Face hallucination based on sparse local-pixel structure," *Pattern Recogn.*, vol. 47, no. 3, pp. 1261–1270, 2014.

[19] Z. Wang, R. Hu, S. Wang, and J. Jiang, "Face hallucination via weighted adaptive sparse regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 802–813, May 2014.

[20] Z. Hui, W. Liu, and K.-M. Lam, "A novel correspondence-based face-hallucination method," *Image and Vision Computing*, 2017.

[21] Y. Hu, N. Wang, D. Tao, X. Gao, and X. Li, "Serf: A simple, effective, robust, and fast image super-resolver from cascaded linear regression," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4091–4102, 2016.

[22] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *IEEE Signal Proc. Let.*, vol. 18, no. 6, pp. 367–370, 2011.

[23] G. Gao and J. Yang, "A novel sparse representation based framework for face image super-resolution," *Neurocomputing*, vol. 134, pp. 92–99, 2014.

[24] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3991–4002, Nov 2017.

[25] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *NIPS*, 2009, pp. 2223–2231.

[26] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *CVPR*, 2010.

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

[28] R. A. Farrugia and C. Guillemot, "Face hallucination using linear models of coupled sparse support," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4562–4577, Sept 2017.

[29] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1674–1684, 2016.

[30] J. Shi, X. Liu, Y. Zong, C. Qi, and G. Zhao, "Hallucinating face image by regularization models in high-resolution feature space," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2980–2995, 2018.

[31] G. Gao, X.-Y. Jing, P. Huang, Q. Zhou, S. Wu, and D. Yue, "Locality-constrained double low-rank representation for effective face hallucination," *IEEE Access*, vol. 4, pp. 8775–8786, 2016.

[32] T. Lu, Z. Xiong, Y. Zhang, B. Wang, and T. Lu, "Robust face super-resolution via locality-constrained low-rank representation," *IEEE Access*, vol. 5, pp. 13 103–13 117, 2017.

[33] 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 Trans. Multimedia*, vol. 19, no. 1, pp. 27–40, 2017.

[34] X. Pei, Y. Guan, P. Cai, and T. Dong, "Face hallucination via gradient constrained sparse representation," *IEEE Access*, vol. 6, pp. 4577–4586, 2018.

[35] L. Liu, C. L. P. Chen, S. Li, Y. Y. Tang, and L. Chen, "Robust face hallucination via locality-constrained bi-layer representation," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1189–1201, 2018.

[36] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902 – 913, 2010.

[37] J. Sun, J. Zhu, and M. F. Tappen, "Context-constrained hallucination for image super-resolution," in *CVPR*, June 2010, pp. 231–238.

[38] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *CVPR*, June 2011, pp. 2833–2840.

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

[40] Y. Hao and C. Qi, "A unified regularization framework for virtual frontal face image synthesis," *IEEE Signal Proc. Let.*, vol. 22, no. 5, pp. 559–563, 2015.

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

[42] J. Shi, X. Liu, and C. Qi, "Global consistency, local sparsity and pixel correlation: A unified framework for face hallucination," *Pattern Recogn.*, vol. 47, no. 11, pp. 3520–3534, 2014.

[43] J. Shi and C. Qi, "Kernel-based face hallucination via dual regularization priors," *IEEE Signal Proc. Let.*, vol. 22, no. 8, pp. 1189–1193, 2015.

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

[45] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3194–3207, 2016.

[46] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *ECCV*. Springer, 2016, pp. 318–333.

[47] S. Zhu, S. Liu, C. C. Loy, and X. Tang, "Deep cascaded bi-network for face hallucination," in *ECCV*. Springer, 2016, pp. 614–630.

[48] Y. Song, J. Zhang, S. He, L. Bao, and Q. Yang, "Learning to hallucinate face images via component generation and enhancement," in *IJCAI*, 2017, pp. 4537–4543.

[49] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li, "Attention-aware face hallucination via deep reinforcement learning," in *CVPR*, 2017, pp. 1656–1664.

[50] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors," in *CVPR*, 2008, pp. 1–8.

[51] X. Yu and F. Porikli, "Imagining the unimaginable faces by deconvolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2747–2761, June 2018.

[52] J. Jiang, Y. Yu, S. Tang, J. Ma, J. Qi, and A. Aizawa, "Context-patch based face hallucination via thresholding locality-constrained representation and reproducing learning," in *ICME*, 2017, pp. 469–474.

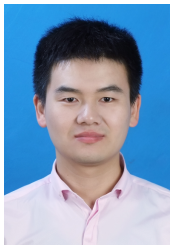
[53] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326.

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

[55] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, feb. 2006.

[56] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch-photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1364–1376, 2013.

[57] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
[58] M. Aharon, M. Elad, and A. Bruckstein, "rmK-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Proc.*, vol. 54, no. 11, pp. 4311–4322, 2006.
[59] X. Liu, D. Zhai, J. Zhou, X. Zhang, D. Zhao, and W. Gao, "Compressive sampling-based image coding for resource-deficient visual communication," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2844–2855, 2016.
[60] X. Liu, D. Zhai, J. Zhou, S. Wang, D. Zhao, and H. Gao, "Sparsity-based image error concealment via adaptive dual dictionary learning and regularization," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 782–796, 2017.
[61] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, 2013.
[62] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–13, 2018.
[63] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol. 45, pp. 153 – 178, 2019.
[64] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multi-layer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4220–4231, 2014.
[65] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *BMVC*, 2012, pp. 1–10.
[66] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *ICCV*, 2015, pp. 370–378.
[67] M. Everingham, J. Sivic, and A. Zisserman, "Hello! My name is... Buffy – Automatic Naming of Characters in TV Video," in *BMVC*, 2006.



Junjun Jiang received the B.S. degree from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014.

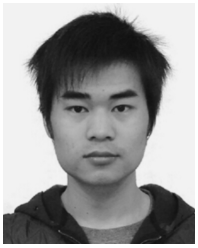
From 2015 to 2018, he was an Associate Professor at China University of Geosciences, Wuhan. Since 2016, he has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. He won the Finalist of the World's FIRST 10K Best Paper Award at ICME 2017, and the Best Student Paper Runner-up Award at MMM 2015. He received the 2016 China Computer Federation (CCF) Outstanding Doctoral Dissertation Award and 2015 ACM Wuhan Doctoral Dissertation Award. His research interests include image processing and computer vision.

Yu Yi is currently an assistant professor with National Institute of Informatics (NII), Japan. Before joining NII, she was a senior research fellow with School of Computing, National University of Singapore. Her research covers large-scale multimedia data mining and pattern analysis, location-based mobile media service and social media analysis. Yu received a Ph.D. in Information and Computer Science from Nara Womens University, Japan.

Suhua Tang received the B.S. degree in electronic engineering and the Ph.D. degree in information and communication engineering from the University of Science and Technology of China, Hefei, China, in 1998 and 2003, respectively.

From October 2003 to March 2014, he was with Adaptive Communications Research Laboratories, Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan. Since April 2014, he has been with the Department of Communication Engineering and Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan, and is a Guest Researcher with ATR. His research interests include green communications, network coding, cross-layer design, mobile ad hoc networks, and intervehicle communications.

Dr. Tang is a member of the Institute of Electronics, Information, and Communication Engineers.



Jiayi Ma received the B.S. degree from the Department of Mathematics, and the Ph.D. Degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was with the Department of Statistics, University of California at Los Angeles. He is now an Associate Professor with the Electronic Information School, Wuhan University. His current research interests include in the areas of computer vision, machine learning, and pattern recognition.

Akiko Aizawa graduated from the Department of Electronics at the University of Tokyo in 1985 and completed her doctoral studies in electrical engineering in 1990. She was a visiting researcher at the University of Illinois at Urbana- Champaign from 1990 to 1992. She is currently a professor in Digital Content and Media Sciences Research Division at National Institute of Informatics. She is also an adjunct professor at the Graduate School of Information Science and Technology at the University of Tokyo. Her research interests include statistical text processing, linguistic resources construction, and corpus-based knowledge acquisition.

Kiyoharu Aizawa received the B.E., the M.E., and the Dr.Eng. degrees in Electrical Engineering all from the University of Tokyo, in 1983, 1985, 1988, respectively. He is currently a Professor at Department of Information and Communication Engineering of the University of Tokyo. He was a Visiting Assistant Professor at University of Illinois from 1990 to 1992. His research interest is in image processing and multimedia applications. He received the 1987 Young Engineer Award and the 1990, 1998 Best Paper Awards, the 1991 Achievement Award, 1999 Electronics Society Award from IEICE Japan, and the 1998 Fujio Frontier Award, the 2002 and 2009 Best Paper Award, and 2013 Achievement award from ITE Japan. He received the IBM Japan Science Prize in 2002. He is currently a Senior Associate Editor of IEEE Tras. Image Processing, and on Editorial Board of ACM TOMM, APSIPA Transactions on Signal and Information Processing, and International Journal of Multimedia Information Retrieval. He served as the Editor in Chief of Journal of ITE Japan, an Associate Editor of IEEE Trans. Image Processing, IEEE Trans. CSVT and IEEE Trans. Multimedia. He has served a number of international and domestic conferences; he was a General co-Chair of ACM Multimedia 2012. He is a council member of Science Council of Japan.