

Big Data Processing With Application to Image Super-Resolution

Xiangjun Meng & Baiqing Diao

State Grid Shandong Electric Power Company, Jinan, China

Lipeng Zhu

China Electric Power Research Institute, Nanjing, China

Guangwei Gao & Song Deng

Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China

ABSTRACT: Learning based face hallucination methods have received much attention in recent years. As opposed to the existing methods, where the input image (patch) matrix is first converted into vectors before combination coefficients calculation, this paper proposes a novel matrix based regression model for directly combination coefficients calculation to preserve the structural information of the input matrix. For each low-resolution local patch matrix, its combination coefficients over the same position image patch matrices in training images can be computed. Then the corresponding high-resolution patch matrix can be obtained. Experiments conducted on the FERET face dataset indicate that our method could outperform other state-of-the-art algorithms in terms of both vision and quantity.

KEYWORD: Face hallucination; Position-patch; Matrix regression

1 INTRODUCTION

Face hallucination, which is also referred as face super-resolution (SR), is a technology to generate high-resolution (HR) face images from low-resolution (LR) inputs. The HR facial images are very useful for video surveillance, 3D facial modeling, automated face recognition, and so on [1].

In recent years, learning based face hallucination methods have received a lot of attentions since they can utilize additional and relevant information from training samples, leading to better super-resolved results than that of reconstruction based ones [2-12]. How to utilize the training set is crucial in learning based methods. For example, Freeman *et al.* [4] utilized a Markov network to model the relationship between LR and HR patches to perform SR. Chang *et al.* [5] utilized the locally linear embedding (LLE) approach and proposed the neighbor embedding (NE) method. NE assumes that the two manifolds constructed by the LR and HR patches respectively have similar structures. Li *et al.* [6] proposed to project the LR and HR patch pairs from the original manifolds into a common manifold with a manifold regularization procedure.

Wang *et al.* [8] suggested the use of principal component analysis (PCA) [9] to represent the structural similarity of face images. PCA is used to represent an input LR image as a linear combination of the LR training samples. The target HR image is obtained by using the same combination coefficients,

but with the LR training samples replaced by the corresponding ones. However, the HR images usually contain ghost artifacts as a result of using global linear constraints on HR images and the adopted PCA-based holistic appearance model. Recently, a novel face hallucination method based on position-patch has been proposed by replacing the global linear constraints with multiple local constraints learned from training patches [10]. Accordingly, the characteristics of the input image can be preserved since the input patch is approximated by position patches. Yang *et al.* [11] used a coupled dictionary learned from all training patches using sparse representation [13] for image SR. They assumed that there exist coupled dictionaries of LR and HR images, which have the same sparse representation for each pair of LR and HR patches. After learning the coupled dictionary pair, the LR image patch is first coded over the LR dictionary by sparse representation. Then the HR patch is reconstructed on HR dictionary with these sparse coefficients.

The previous methods, such as [5, 8, 10, 11], are all vector-based method. That is to say, before computing the combination coefficients of the input LR image (patch) matrix over the training samples, we have to convert the images (patches) matrix into vectors in advance. However, in the converting step, some structural information (e.g. the rank of matrix) might be lost. The resulting HR images may contain some noisy artifacts along contours.

In this paper, we will develop a novel matrix based regression model, which can compute the combination coefficients straightforward (without the matrix-to-vector conversion). We will make full use of the information of image matrices and use the minimal rank of representation residual matrix as a criterion to determine the combination coefficients. Inspired by the work in [10], we also use the position-patch instead of neighbor-patch in representation and reconstruction. First, each input LR image patch matrix is directly represented as a linear combination of the same position image patch matrices in each training image using nuclear norm based matrix regression. Then, the HR image patch matrix is reconstructed based on the corresponding HR training patch matrix with the same combination coefficients. The whole HR image can be synthesized by integrating the reconstructed patches. The performance of the proposed method is examined and compared with other state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 formulates the nuclear norm based matrix regression. Section 3 presents our face hallucination algorithm. Section 4 reports the experimental results and Section 5 concludes the paper.

2 NUCLEAR NORM BASED CODING

In this section, we describe the nuclear norm based matrix regression and then use the alternating direction method of multipliers to solve the problem

2.1 Formulation

Given a set of n observed matrices $A_1, \dots, A_n \in \mathbb{R}^{p \times q}$ and a query data matrix $D \in \mathbb{R}^{p \times q}$, D can be represented linearly using A_1, \dots, A_n , i.e., $D = x_1 A_1 + x_2 A_2 + \dots + x_n A_n + E$, where x_1, \dots, x_n is a set of representation coefficients and E is the representation residual. Let us denote the following mapping:

$$A(x) = x_1 A_1 + x_2 A_2 + \dots + x_n A_n \quad (1)$$

Then, we can have $E = D - A(x)$. The nuclear norm based coding can be formulated as follows [14]:

$$\min_{x, E} \|x\|_1 + \lambda \|E\|_*, \quad s.t. \quad A(x) - D = E \quad (2)$$

where $\|\cdot\|_1$ denotes the l_1 -norm of a vector, $\|\cdot\|_*$ denotes the nuclear norm of a matrix

For the convenience of expression, (1) can be converted into the following equivalent problem:

$$\min_{x, z, E} \|z\|_1 + \lambda \|E\|_*, \quad s.t. \quad A(x) - D = E \text{ and } x = z. \quad (3)$$

2.2 Solution

The alternating direction method of multipliers (ADMM) or the augmented Lagrange multipliers (ALM) method has been applied to solve the nuclear norm optimization problems [15]. Here, we give the process of using ADMM to solve the regularized matrix regression problem.

The augmented Lagrangian function of Eq. (3) is

$$L_\mu = \|z\|_1 + \lambda \|E\|_* + \text{tr}(y_1^T (x - z)) + \text{tr}(Y_2^T (A(x) - D - E)) + \frac{\mu}{2} (\|x - z\|_F^2 + \|A(x) - D - E\|_F^2), \quad (4)$$

where $\mu > 0$ is a penalty parameter, y_1 and Y_2 are the Lagrange multipliers, $\text{tr}(\cdot)$ is the trace operator. Following some simple algebraic steps, Eq. (4) can be rewritten as

$$L_\mu = \|z\|_1 + \lambda \|E\|_* + \frac{\mu}{2} \left(\left\| x - z + \frac{1}{\mu} y_1 \right\|_F^2 + \left\| A(x) - D - E + \frac{1}{\mu} Y_2 \right\|_F^2 \right) - \frac{1}{2\mu} \|y_1\|_2^2 - \frac{1}{2\mu} \|Y_2\|_F^2. \quad (5)$$

The function L can be minimized by updating each of variables one at a time. The detailed algorithm via ADMM is summarized in Algorithm 1 [14]. For the step 1, the soft-thresholding operator [15] can be applied. For the step 2, the singular value thresholding algorithm [16] can be applied.

Algorithm 1[14]. Nuclear norm based coding via ADMM

Input: Training matrices $A_1, \dots, A_n \in \mathbb{R}^{p \times q}$ and a matrix $B \in \mathbb{R}^{p \times q}$, parameters λ and μ , the termination condition parameter ε .

1: Fix the others and update z^{k+1} by

$$z^{k+1} = \arg \min_z \frac{1}{\mu} \|z\|_1 + \frac{1}{2} \left\| z - \left(x^k + \frac{1}{\mu} y_1^k \right) \right\|_F^2;$$

2: Fix the others and update x^{k+1} by

$$x^{k+1} = \arg \min_x \frac{\mu}{2} \left(\left\| x - z^{k+1} + \frac{1}{\mu} y_1^k \right\|_F^2 + \left\| A(x) - B - E^k + \frac{1}{\mu} Y_2^k \right\|_F^2 \right);$$

3: Fix the others and update E^{k+1} by

$$E^{k+1} = \arg \min_E \frac{\lambda}{\mu} \|E\|_* + \frac{1}{2} \left\| E - \left(A(x^{k+1}) - B + \frac{1}{\mu} Y_2^k \right) \right\|_F^2;$$

4: Update the multiplies

$$y_1^{k+1} = y_1^k + \mu(x^{k+1} - z^{k+1}), \quad Y_2^{k+1} = Y_2^k + \mu(A(x^{k+1}) - B - E^{k+1});$$

5. If the termination condition is satisfied, go to 6; otherwise go to 1.

6. **Output:** Optimal regression coefficient vector x^{k+1}

3 FACE IMAGE SUPER-RESOLUTION BASED ON MATRIX REGRESSION

Denote the patches of LR testing image, patches of the m^{th} LR training image, patches of the m^{th} HR training image located at position (i, j) as $\{B_L^p(i, j)\}_{p=1}^N$, $\{A_L^{mp}(i, j)\}_{p=1}^N$ and $\{A_H^{mp}(i, j)\}_{p=1}^N$ respectively. N is the number of patches. Each patch is represented in the form of matrix.

Let us define the following linear mapping:

$$A_L^p(i, j)(x) = \sum_{m=1}^M x_m(i, j) A_L^{mp}(i, j) \quad (6)$$

where M is the number of training images.

For each patch matrix $B_L^p(i, j)$ in the LR testing image, its reconstruction weights can be obtained by

$$\hat{x}(i, j) = \arg \min_{x(i, j)} \|x(i, j)\|_1 + \lambda \|A_L^p(i, j)(x) - B_L^p(i, j)\|_2 \quad (7)$$

Eq. (7) can be solved by Algorithm 1.

Replacing each LR image patch matrix $A_L^{mp}(i, j)$ by its corresponding HR sample $A_H^{mp}(i, j)$, the desired HR patch is denoted as

$$\sum_{m=1}^M \hat{x}_m(i, j) A_H^{mp}(i, j) = B_H^p(i, j) \quad (8)$$

The whole HR image can be synthesized by integrating the reconstructed patches according to the original position. Pixels in the overlapping regions are obtained by averaging the pixels value in the overlapping regions between two adjacent patches reconstructed.

The proposed face hallucination algorithm is summarized in Algorithm 2.

Algorithm 2 Proposed face hallucination algorithm

Input: LR training matrices A_L^1, \dots, A_L^M , HR training matrices A_H^1, \dots, A_H^M , LR input matrices B_L .

- 1: Divide the LR input image and training image into overlapped patch matrices;
 - 2: For each patch $B_L^p(i, j)$:
 - (a) Compute the reconstruct weights $x(i, j)$ by Eq. (7);
 - (b) Synthesize the HR patch $B_H^p(i, j)$ by Eq. (8).
 - 3: Integrate the HR patches to form target HR image B_H .
-

4. Output: The Synthesized HR image B_H

4 EXPERIMENTAL RESULTS

To verify the superiority of our method, experiments were conducted on the FERET database [17]. We selected 1196 face images of different people in our experiments. These face images were aligned manually using the locations of three points: centers of left and right eyeballs and center of the mouth. All faces are standardized to the size of 120×102 .

These HR images were blurred and down-samples to generate LR images with size 40×34 . We compared our approach with bicubic interpolation, Wang's method [8], Chang's method [5], Ma's method [10] and Yang's method [11]. In order to get the optimal results in Wang's method, we used image pairs of 1176 people for training and let variance contribution rate of PCA be 0.90. The number of neighbor-patches in Chang's method is 380. Image pairs of 380 people were used for training in Chang's method, Ma's method, Yang's method and our method. The other images of 20 people were used as test images. The LR image patch size was 3×3 and the patch overlapped with its adjacent patch by 3 pixels (i.e., one column). The parameters λ and μ in our method are set to be 0.01 and 0.1.

Fig. 1 show some hallucinated results visually. The results of Wang's method can hardly maintain global smoothness around the face contour and margin of the mouth. The results of Chang's method are somewhat blurred. There are jaggy artifacts along the eyes in Yang's method. While Ma's method and our method produce more nature looking images than them. To quantify the results from the ground truth data, Table 1 tabulates the average PSNR, SSIM [18] and FSIM [19] of the reconstructed facial images using respective methods. From the table, we can see that our method with nonlocal self-similarities achieves the best hallucination performance.

Table 1. Average PSNR, SSIM and FSIM for each method

Quality measure	PSNR	SSIM	FSIM
Wang's method [8]	27.3729	0.7521	0.8700
Chang's method [5]	29.5950	0.8795	0.9152
Yang's method [11]	29.7871	0.8704	0.9163
Ma's method [10]	30.7363	0.8883	0.9236
Our method	30.8514	0.8935	0.9280

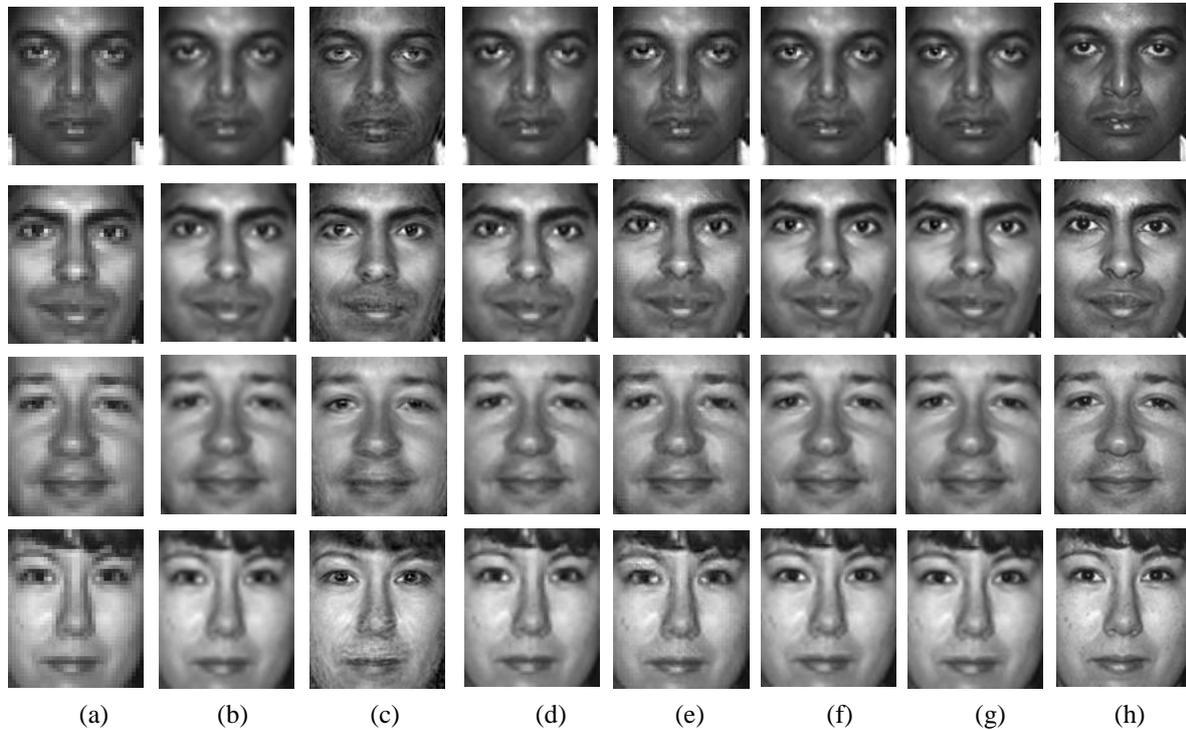


Figure 1: The hallucination results. (a) Input 40×34 LR faces. (b) Bicubic interpolation. (c) Wang's method [8]. (d) Chang's method [5]. (e) Yang's method [11]. (f) Ma's method [10]. (g) Our method. (h) The original 120×102 HR faces.

5 CONCLUSION

In this paper, we proposed a novel matrix based regression with local position-patch for efficient face hallucination. The proposed method can compute combination coefficients of input low-resolution patch matrix directly without matrix-to-vector conversion. Experimental results on FERET face dataset indicate that our method could yield better reconstruction results than other state-of-the-art methods.

REFERENCES

- [1] Baker, S. & Kanade, T. 2002. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(9): 1167-1183.
- [2] Dai, S. et al. 2007. Soft edge smoothness prior for alpha channel super resolution. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, 17-22 June 2007*.
- [3] Dong, W. et al. 2009. Nonlocal back-projection for adaptive image enlargement. *IEEE International Conference on Image Processing, Cairo, 7-10 Nov. 2009*.
- [4] Freeman, W. et al. 2002. Example-based super-resolution. *IEEE Computer Graphics and Applications* 22(2): 56- 65.
- [5] Chang, H. et al. 2004. Super-resolution through neighbor embedding. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, 27 June- 2 July 2004*.
- [6] Li, B. et al. Aligning coupled manifolds for face hallucination. *IEEE Signal Processing Letters* 16(11): 957- 960.
- [7] Par, J. S. & Lee, S. W. 2008. An example-based face hallucination method for single-frame, low-resolution facial images. *IEEE Transactions on Image Processing* 17(20): 1806- 1816.
- [8] Wang, X. & Tang, X. 2005. Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Application and Reviews* 35(3): 425- 434.
- [9] Jain, A.K. et al. 2000. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1): 4- 37.
- [10] Ma, X. et al. 2010. Hallucinating face by position-patch. *Pattern Recognition* 43(6): 2224- 2236.
- [11] Yang, J. et al. 2010. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* 19(11): 2861- 2873.
- [12] Zhang, W. & Cham, W. K. 2008. Learning-based face hallucination in DCT domain. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, 23-28 June 2008*.
- [13] Wright, J. et al. 2009. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2): 210- 227.
- [14] Luo, L. et al. 2014. Nuclear Norm Regularized Sparse Coding. *IEEE International Conference on Image Processing*.
- [15] Lin, Z. et al. 2009. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *UIUC Tech. Rep. UIUC-ENG-09-2215*.
- [16] Cai, J.F. et al. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4): 1956- 1982.
- [17] Philips, P. et al. 1997. The FERET evaluation methodology for face recognition algorithms. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, 17-19 June 1997*.
- [18] Wang, Z. & Bovik, A.C. 2009. Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine* 26(1): 98- 117.
- [19] Zhang, L. et al. 2011. FSIM: a feature similarity index for image quality assessment. *IEEE Transactions on Image Processing* 20(8): 2378- 2386.