FACE HALLUCINATION VIA SEMI-KERNEL PARTIAL LEAST SQUARES

Qiang Zhang, Fei Zhou, Fan Yang and Qingmin Liao,IEEE Member

Shenzhen Key Lab of Information Sci. &Tech., Shenzhen Engineering Lab of IS &DRM, Department of Electronics Engineering / the Graduate School at Shenzhen, Tsinghua University, China

zqtsinghua@126.com; flying.zhou@163.com; fyang@sz.tsinghua.edu.cn; liaoqm@sz.tsinghua.edu.cn

Keywords-face super-resolution (SR); collaborative representation (CR); semi-kernel partial least squares (semi-KPLS);

Abstract—In this paper, a patch-based super-resolution (SR) method is proposed to hallucinate facial images. Two steps are involved in this method. In the first step, semi-kernel partial least squares (semi-KPLS) algorithm is used to generate a nonlinear correlative space. In the second step, we use collaborative representation (CR) to infer a high-resolution (HR) face. The experiments conducted on FERET database demonstrate the proposed algorithm can generate better results, in comparison with some state-of-the-art approaches.

Introduction

The details of facial features are crucial for identifying an individual from surveillance video, but the resolution of a facial region from surveillance videos is usually very low. Therefore, in order to obtain detailed facial features, it is necessary to infer a high-resolution (HR) face image from the low-resolution (LR) one. This technique is called face hallucination or face super-resolution (SR).

According to the available inputs, existing SR methods can roughly be divided into two classes: multi-frames SR [1-4], and single-frame SR [5-11]. The inspiration of multi-frames SR is to retrieve high-frequency details from complementary multi-images. This type of method suffers from ill-conditioned registration and is limited to small increase in image spatial resolution [15]. Instead, single-frame SR prefers to infer the HR counterpart of a single image by obtaining extra information from training samples and we mainly focus on this approach in this paper. In [14], Freeman et al. propose an example-based method using a Markov Random Field (MRF) with belief propagation, and HR images are generated by building a one-to-one relation between HR and LR patches in training samples. In [18], based on the principle of locally linear embedding (LLE), HR patches are constructed by a linear combination of several neighbors. However, most SR methods are developed for the purpose of super-resolving general images.

For face SR, the utilization of the properties of face is conductive to generate the high-resolution face images. Baker et al. [19] were the first to develop a hallucination method under a Bayesian formulation and proposed the term “face hallucination”. In this method, it generates the high frequency details from a parent structure with the assistance of training samples. This work is further extended by Su et al. [23], where a steerable pyramid is used to provide larger search scope for local best match approach. Eigen-transformation is introduced to hallucinating face in [5], whose conclusion results in the example-based paradigm. That is hallucinated face can be represented by a linear combination of training samples. In [24], canonical correlation analysis is employed to maximize the correlation coefficients between LR and HR data. However, subspace transformations may reduce the discrimination of hallucinated faces. Liu et al. [25] developed a two-step approach integrating a global parametric model with Gaussian assumption and a local non-parametric model based on Markov random field (MRF). Zhuang et al. [6] use locality preserving projection and neighbor embedding to hallucinate a high-resolution face. More recently, the work in [8] shows that patch-based method can achieve plausible results without the residue compensation. In fact, the residue compensation is merely indispensable for the methods that include the step of dimension reduction and HR patches are inferred using the training patches of the same position. Instead of performing face SR in spatial domain, the SR method in discrete cosine transform (DCT) is developed in [9]. Wu et al. [10] use kernel partial least squares (KPLS) to obtain the high-resolution image which do not consider the high-frequency details. In [11], multilateral affinity function (MAF) is utilized to select and weight image patches which are employed to reconstruct HR image. However, all these methods cannot solve the blurring issue.

In this paper, a new method for solving face image SR problem is proposed. Inspired by KPLS [10] and position-patch based face hallucination approach [8], we first code the LR image patches in a nonlinear correlative space and recover HR patch with the coding coefficients. In [10], the regression analysis of KPLS is used to infer the HR image, which cannot obtain high-frequency details. Unlike [10], we just take advantage of the KPLS to generate nonlinear correlative space. That is why the proposed method is called semi-KPLS. The true that compared with sparse representation, CR has very competitive face recognition accuracy but with significantly lower complexity has been proven in [13]. That is why CR is used to code the LR image patches in the nonlinear correlative space. The coding coefficients are used to generate correspond HR image patches, and the final HR image can be obtained by merging these HR image patches.

The rest of the paper is organized as follows. Section Ⅱ presents the details about the proposed method. Experimental results and discussions are described in Section Ⅲ. Section Ⅳ concludes this paper.

FACE HALLUCIANTION METHOD

A. Collaborative representation(CR)

In [7], image SR is approached via sparse representation of raw image patches. Image patches can be coded sparsely with respect to a dictionary, i.e. $\mathbf{x} = \mathbf{D}_h\boldsymbol{\alpha}$, $\mathbf{y} = \mathbf{D}_l\boldsymbol{\alpha}$ with $\|\boldsymbol{\alpha}\|_0 \ll N$ where $\mathbf{x}$ is HR image patch and $\mathbf{y}$ is the corresponding LR input, $\mathbf{D}_h = [\mathbf{x}_1, ..., \mathbf{x}_N] \in \mathbb{R}^{d_h \times N}$ and $\mathbf{D}_l = [\mathbf{y}_1, ..., \mathbf{y}_N] \in \mathbb{R}^{d_l \times N}$ are the HR dictionary and the corresponding LR dictionary. $\mathbf{D}_l$ is used to train the coding vector and $\mathbf{D}_h$ is utilized to reconstruct HR image patch. N is the patch number of the dictionary. $\|\cdot\|_0$ denotes L_0 norm. $\boldsymbol{\alpha}$ is the coding vector. Note that the coding coefficients of HR and LR patch pairs are the same. Actually, the coding coefficients may not be strictly equal without considering the correlation between $\mathbf{D}_l$ and $\mathbf{D}_h$, and we will solve this problem in Section B. When reconstructing an HR patch $\mathbf{x}$ from LR patch $\mathbf{y}$, the sparse coefficient vector $\boldsymbol{\alpha}$ can be estimated as:

$$\hat{\boldsymbol{\alpha}} = arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}_l\boldsymbol{\alpha}\|_2^2 + \gamma \|\boldsymbol{\alpha}\|_1 \quad (1)$$

Then, HR image patch is reconstructed by the coefficients calculated in (1): $\mathbf{x} = \mathbf{D}_h\hat{\boldsymbol{\alpha}}$

Since human face is highly structured, position feature is important for reconstruction [11]. Inspired by [8], we adopt position patches instead of randomly selected patches for training. Position-patches are defined as the patches in all the training images that have the same position with respect to a certain patch in the LR input. An LR or HR image is represented by a set of image patches that overlap with each other. Incorporating the position feature, we design dictionaries for HR and LR image patch at position (i, j) as: $\mathbf{D}_h(i,j) = [\mathbf{x}_1(i,j), ..., \mathbf{x}_N(i,j)]$, $\mathbf{D}_l(i,j) = [\mathbf{y}_1(i,j), ..., \mathbf{y}_N(i,j)]$ where $\mathbf{x}_k(i,j)$, and $\mathbf{y}_k(i,j)$ are patch pairs at position (i,j) from training images, which include vector of pixel values. Zhang et al. revealed that the collaborative representation (CR) mechanism, instead of the l_1-norm sparsity constraint (sparse representation), which truly improves the face recognition accuracy [13]. With significantly low complexity and competitive face recognition accuracy, the CR is used to code the dictionaries.

$$\hat{\boldsymbol{\alpha}} = arg\min_{\boldsymbol{\alpha}} \left\|\mathbf{y}(i,j) - \mathbf{D}_l(i,j)\boldsymbol{\alpha}\right\|_2^2 + \gamma\left\|\boldsymbol{\alpha}\right\|_2^2 \quad (2)$$

where γ is a regularized scalar. Then, the inferred HR patch is calculated as: $\mathbf{x}(i,j) = \mathbf{D}_h(i,j)\hat{\boldsymbol{\alpha}}$. For convenience, position (i,j) is omitted in the following, i.e. $\mathbf{y}$ represents $\mathbf{y}(i,j)$.

B. Coding in nonlinear correlative space

Recent years, the kernel trick is a well-known technique in machine learning, which can generalize a linear algorithm to its nonlinear counterpart. The key idea of the kernel approach is to map the input data to a high-dimensional feature space Γ corresponding to a reproducing kernel Hilbert space, where the nonlinear structure in the input space is more likely to be linear. The nonlinear mapping can be described as $\Phi : \mathbb{R}^d \to \mathbb{R}^D$, $D \gg d$. $\Phi(\mathbf{t}) \in \mathbb{R}^D$ is the image of t in the feature space Γ. Gaussian kernel of width $\sigma > 0$ is introduced:

$$\mathbf{K}(\mathbf{y}_m, \mathbf{y}_n) = \Phi(\mathbf{y}_m)^T \Phi(\mathbf{y}_n) = \exp(-\left\|\mathbf{y}_m - \mathbf{y}_n\right\|^2 / 2\sigma^2) \quad (3)$$

Inspired by [10] and the principle of kernel trick, the objective function of semi-KPLS (without regression analysis):

$$\max_{|\mathbf{r}|=|\mathbf{s}|=1} \left\{ var(\Phi_{\mathbf{D}_l}\mathbf{r})\left[corr(\Phi_{\mathbf{D}_l}\mathbf{r}, \mathbf{D}_h\mathbf{s})\right]^2 var(\mathbf{D}_h\mathbf{s}) \right\} \quad (4)$$

where var, and $corr$ denote the sample variance and squared correlation respectively. The $\mathbf{r}$ and $\mathbf{s}$ denote the weight vectors. $|\cdot|$ is used to obtain the length of vector. $\Phi_{\mathbf{D}_l} = \Phi_{\mathbf{D}_l}(i,j) = \left[\Phi(\mathbf{y}_1(i,j)), ..., \Phi(\mathbf{y}_N(i,j))\right] \in \mathbb{R}^{D\times N}$ is the LR dictionary in Γ. Equation (4) denotes that the semi-KPLS algorithm balances with the requirement to explain as much variance as possible in both $\Phi_{\mathbf{D}_l}$ and $\mathbf{D}_h$ by using the criterion of maximal correlation. Therefore, in the nonlinear correlative space $\Phi(\mathbf{D}_l)\mathbf{r}$, the coding vector $\boldsymbol{\alpha}$ of the LR patch can be used to reconstruct correspond HR patch directly due to the correlation. Equation (4) can be simplified to:

$$\Phi_{\mathbf{D}_l}\mathbf{D}_h^T\mathbf{D}_h\Phi_{\mathbf{D}_l}^T\mathbf{r} = \lambda\mathbf{r} \quad (5)$$

When $\mathbf{r}$ is the eigenvector correspond to the maximum eigenvalue in $\boldsymbol{\lambda}$, the maximum value in (4) is achieved. It is not practical to directly solve (5) to get the projection matrix which is formed by eigenvectors correspond to the top M eigenvalues, because the mapping $\Phi(\cdot)$ is usually unknown. We multiply $\Phi_{\mathbf{D}_l}^T$ by the two sides of equation (5) to get the new feature dictionary.

$$\Phi_{\mathbf{D}_l}^T\Phi_{\mathbf{D}_l}\mathbf{D}_h^T\mathbf{D}_h[\Phi_{\mathbf{D}_l}^T\mathbf{r}] = \lambda[\Phi_{\mathbf{D}_l}^T\mathbf{r}] \quad (6)$$

Based on (6) and $\mathbf{K} = \Phi_{\mathbf{D}_l}^T\Phi_{\mathbf{D}_l} = \mathbf{K}(\mathbf{y}_i, \mathbf{y}_j)$, the eigenvectors correspond to the top M eigenvalues of matrix $\mathbf{K}\mathbf{D}_h^T\mathbf{D}_h$ are used to form a matrix $\mathbf{Q} \in \mathbb{R}^{N\times M}$. The new feature dictionary is:

$$\mathbf{D}_{new-l} = \mathbf{D}_{new-l}(i,j) = [\mathbf{p}_1{}^T(i,j), ..., \mathbf{p}_N{}^T(i,j)] \in \mathbb{R}^{M\times N}$$

where $\mathbf{p}_k(i,j) \in \mathbb{R}^{1\times M}$ is the k-th row of $\mathbf{Q}$. To get the new feature of $\mathbf{y}$, $\Phi^T(\mathbf{y})$ is multiplied by the two sides of (5).

$$\Phi^T(\mathbf{y})\Phi_{\mathbf{D}_l}\mathbf{D}_h^T\mathbf{D}_h[\Phi_{\mathbf{D}_l}^T\mathbf{r}] = \lambda\Phi^T(\mathbf{y})\mathbf{r} \quad (7)$$

$$\Phi^T(\mathbf{y})\Phi_{\mathbf{D}_l} = \mathbf{K}(\cdot, \mathbf{y}) = \left[\mathbf{K}(\mathbf{y}_1, \mathbf{y}), \ldots \mathbf{K}(\mathbf{y}_N, \mathbf{y})\right]^T \quad (8)$$

From (7) and (8), we can obtain the new feature of y as:

$$\mathbf{y}_{new} = \mathbf{y}_{new}(i,j) = \Phi^T(\mathbf{y})\mathbf{r} = \mathbf{K}(\cdot, \mathbf{y})\mathbf{D}_h^T\mathbf{D}_h[\Phi_{\mathbf{D}_l}^T\mathbf{r}] / \lambda \quad (9)$$

where $\Phi_{\mathbf{D}_l}^T\mathbf{r}$ is obtained from (6). Substituting $\mathbf{y}_{new}$, $\mathbf{D}_{new-l}$ into (2), we get:

$$\hat{\boldsymbol{\alpha}} = arg\min_{\boldsymbol{\alpha}} \left\|\mathbf{y}_{new}(i,j) - \mathbf{D}_{new-l}(i,j)\boldsymbol{\alpha}\right\|_2^2 + \gamma\left\|\boldsymbol{\alpha}\right\|_2^2$$

The solution can be easily determined as: $\hat{\boldsymbol{\alpha}} = (\mathbf{D}_{new-l}{}^T\mathbf{D}_{new-l} + \gamma \cdot \mathbf{I})^{-1}\mathbf{D}_{new-l}{}^T\mathbf{y}_{new}$

$\mathbf{I}$ is identity matrix. Then, we can reconstruct its HR counterpart as: $\mathbf{x} = \mathbf{D}_h \boldsymbol{\alpha}$. All the HR patches $\mathbf{x}(i, j)$ are integrated to form the global HR image according to the position of patches. Pixels of the overlapping regions in the result are obtained by averaging the pixel values in the overlapping regions between two adjacent patches hallucinated. The flow chart of the proposed hallucination method is shown in Fig. 1.

EXPERIMENTS

C. Experimental settings

The proposed algorithm is conducted on the FERET database [22]. 300 face images are randomly selected as training set, and all the images are aligned manually using the three points: centers of eyeballs and mouth. The region of faces are cut out and standardized to the size of 126×126 pixels. For the HR reconstruction, zooming factor q is 7. Consequently, LR image is of size 18×18, and the size of LR patch is 3×3, thus HR patch size is 21×21. An overlap of 1 pixel between adjacent LR patches is adopted, so 7 pixels overlap between adjacent HR patches. Inspired by [8], the training patches for a given LR patch are selected in the same position. Other parameters settings are as follows: σ is set to 100 in (3), $\gamma = 1.0 \times 10^{-6}$, and reduced dimensionality M is set to 100.

D. Result and discussions

To demonstrate the effectiveness of our method, we compare it with classical methods including the Eigen Transformation (ET) [5], the LPH [6], the Position-Patch (PP) [8], the DCT Patch (DP) [9], and the MAF [11] and the results are shown in Fig.2. From these illustrations, our method can generate sharper results with fewer artifacts than the competitive. Root-mean-square error (RMSE), feature-similarity (FSIM) [17], and structural similarity (SSIM) [16] are employed to measure the face hallucination performance. Better SR results should have lower RMSE, higher SSIM, and FSIM. The quantitative comparisons are provided in Table 1, where the values are averaged over 300 of the testing images. From Table 1, we can find that our method outperforms the compared methods in terms of various criteria of signal fidelity. Both representatives global-based and local-based face recognition algorithms are adopted to provide a comprehensive comparison. Uniform pursuit (UP) approach [20] and local binary pattern (LBP) approach [21] are used. The rank-one recognition rates are recorded in Table 2. In the comparisons, the recognition rates of the original HR images are also provided as benchmarks. The results with the recognition rates closer to the benchmarks are believed to be better. From the results, we can find that the proposed algorithm outperforms existing face SR methods for both global and local-based recognition engines.

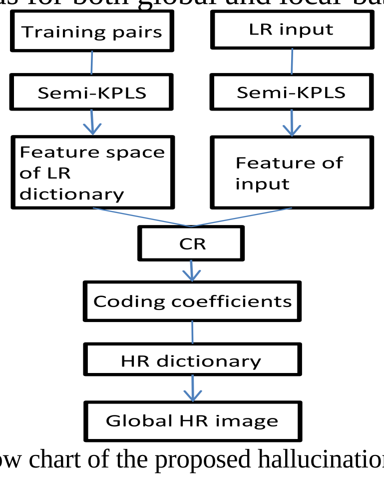

Fig.1: Flow chart of the proposed hallucination method

CONCLUSION

In this paper, we propose a new learning-based face SR method. Inspired by the patch-based position approaches, semi-KPLS is proposed to generate nonlinear correlative space and CR is utilized to obtain the coding coefficients in the new space. Considering the nonlinear similarity of face features and the correlative property, we can retain the global face features in the reconstructed image. Experimental results show that our method outperforms the state-of-art algorithms

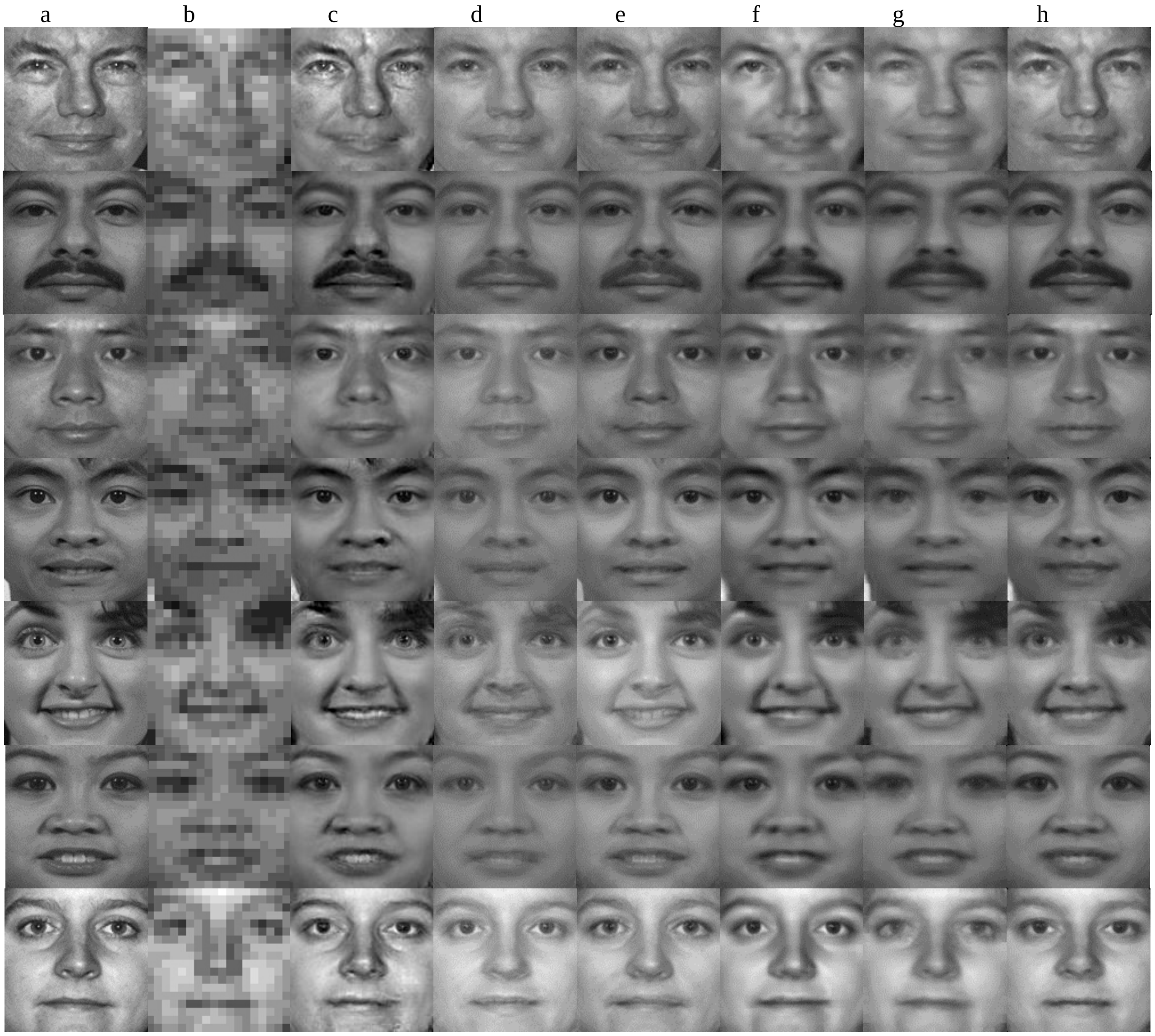

Fig.2: visual comparisons of different methods: (a) HR original images, which serve as the ground true in comparisons, (b) LR inputs. (c) Final results of our method. (d)-(h) Results of ET [5], LPH [6], PP [8], DP[9], and MAF [11], respectively.

Table 1: Quantitative comparisons on signal fidelity

Database	Measurement metric	ET[5]	LPH[6]	PP[8]	DP[9]	MAF[11]	Ours
FERET	RMSE	6.104	5.980	6.528	5.810	5.494	**5.200**
	SSIM	0.7201	0.8075	0.7855	0.8431	0.8713	**0.8901**
	FSIM	0.8453	0.8854	0.8767	0.8895	0.9014	**0.9122**

Table 2: Quantitative comparisons on rank one recognition rates (%)

Database	Recognition method	Original HR (benchmarks)	LR	ET[5]	LPH[6]	PP[8]	DP[9]	MAF[11]	Ours
FERET	UP[20]	92.33	78.33	75.67	60.67	78.00	79.67	85.33	**87.55**
	LBP[21]	96.00	70.81	72.67	73.00	67.00	66.00	91.00	**92.77**

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No.61271393 and 61301183), the China Post-Doctoral Science Foundation (Grant No.

2013M540947 and 2014T70083), and the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen（Grant No. JCYJ20140417115840272).

References

[1] S.C. Park, M.K.Park, and M.G.Kang, "Super-resolution image reconstruction: a technique overview," IEEE Signal Processing Magazine, IEEE, vol.20, no. 3, pp. 21-36, 2003.

[2] A.J.Patti, M.I.Sezan, and A.M.Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," IEEE Transactions on Image Processing, IEEE, vol. 6, no.8 pp. 1064-1076, 1997.

[3] E. Levitan and G.T. Herman, "A maximum posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," IEEE Transactions on Medical Imaging, IEEE, vol. 6, pp. 185–192, 1987.

[4] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multi-surface fitting," IEEE Trans. Image Processing, IEEE, vol. 21, no. 7, pp. 3312-3318, 2012.

[5] X. Wang and X. Tang, "hallucinating face by eigen-transformation," IEEE Transactions on systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE, vol. 35, no. 3, pp.425-434, 2005.

[6] Y. Zhuang, J. Zhang, and F. Wu, "Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation," Pattern Recognition, vol. 40, no.11, pp. 3179-3194, 2007.

[7] J. Yang, J. Wright, T. Huang, and Y. Ma, "image super-resolution as sparse representation of raw image patches," IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 1-8, 2008.

[8] X. Ma, J. Zhang, and C. Qi, "Position- based face hallucination method," IEEE International Conference on Multimedia and Expo, 2009. ICME 2009, IEEE, pp.290-293, 2009.

[9] W. Zhang, W.-K. Cham, "Hallucinating face in the DCT domain," IEEE Transactions Image Processing, IEEE, vol.20, no.10, pp. 2769-2779, 2011.

[10] W. Wu, Z. Liu, and X. He, "Learning-based super resolution using kernel partial least squares," Image and Vision Computing, vol. 29, pp. 394-406, 2011.

[11] F. Zhou, B. Wang, and Q. liao, "Super-resolution for facial image using multilateral affinity function," Neurocomputing, vol. 133, pp. 194-208, 2014.

[12] Q. Liu, S. Wang, L. Ying, X. Peng, Y. Zhu, and D. Liang, "Adaptive dictionary learning in sparse gradient domain for image recovery," IEEE Transactions on Image Processing, IEEE, vol.22, no. 12, pp. 4652-4663, 2013.

[13] D. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, pp. 471-478, 2011.

[14] W.T.Freeman,T.R.Jones,E.Z.Pasztor, "Example-based superresolution," IEEE Comput. Graph.Appl.22(2)(2002)56–65.

[15] S.Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, vol.24, no. 9, pp. 1167-1183, 2002.

[16] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, IEEE, vo. 13, no. 4, pp. 600-612, 2004.

[17] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," IEEE Transactions on Image Processing, IEEE, vol. 20, no. 8, pp. 2378-2386, 2011.

[18] H.Chang, D.-Y.Yeung, Y.M.Xiong, “Super-resolution through neighbor embed-ding,” in:Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,Washington, pp.1275–1282, 2004.

[19] Baker,S. and Kanade, T. “hallucinating faces,” IEEE International Conference on automatic face and gesture recognition, pp.83-88, 2000.

[20] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, “Robust, accurate and efficient face recognition from a single training image: a uniform pursuit approach,” Pattern Recognition, vol. 43, no. 5, pp. 1748-1762, 2010.

[21] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: application to face recognition,” IEEE Transcations on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, 2006.

[22] P.J.Phillips, H.Moon, S.A.Rizvi, and P.J.Rauss, “The FERET evaluation methodology for face recognition algorithms,” IEEE Trans.Pattern Anal. Match. Intell. vol. 22, no. 10, pp. 1090-1104, 2000.

[23] C. Su, Y. Zhuang, L. Huang and F. Wu, “Steerable pyramid-based face hallucination,” Pattern Recognition vo. 38, no. 6, pp. 813-824, 2005.

[24] H. Huang, H. He, X. Fan and J. Zhang, “Super-resolution of human facial image using canonical correlation analysis,” Pattern Recognition, vo. 43, no. 7, pp. 2532-2543, 2010.

[25] C. Liu, H. Shum, and C. Zhang, “A two-step approach to hallucinating faces,” Computer Vision and Pattern Recognition, vo. 1, pp.I-192-I-198,2001