Global Face Reconstruction for Face Hallucination Using Orthogonal Canonical Correlation Analysis

Huiling Zhou[*], Jiwei Hu[†] and Kin-Man Lam[*]
[*] Centre for Signal Processing, Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Kowloon, Hong Kong
E-mail: hl.zhou@connect.polyu.hk, enkmlam@polyu.edu.hk
† Wuhan University of Technology, Wuhan China
E-mail: hujiwei@whut.edu.cn

Abstract— **In this paper, a global face reconstruction framework for face hallucination is proposed to globally reconstruct a high-resolution (HR) version of a face from an input low-resolution (LR) face, based on learning from LR-HR face pairs using orthogonal canonical correlation analysis (orthogonal CCA). In our proposed algorithm, face images are first represented using principal component analysis (PCA). CCA with the orthogonality property is then employed to maximize the correlation between the PCA coefficients of the LR and the HR face pairs so as to improve the hallucination performance. The original CCA does not own the orthogonality property, which is crucial for information reconstruction. In this paper, we utilize an orthogonal variant of CCA, which has been proven by experiments to achieve a better performance than the original CCA in terms of global face reconstruction.**

I. INTRODUCTION

Human face super-resolution (also known as face hallucination) [1] is a recent, hot research topic which aims to generate a high-resolution (HR) face image from one or multiple input low-resolution (LR) face images. It can be applied to various real-life tasks such as video surveillance, face recognition, image-based rendering, etc. In general, face-hallucination techniques can be divided into two categories: reconstruction-based methods [2-3] and learning-based methods [4-10]. The first group of methods reconstructs a HR image based on its LR counterpart only, without any referencing to additional, external information. The learning-based methods explore the correlation between a set of HR face images and their corresponding LR counterparts in the reconstruction of the HR version of a LR face image. Due to the fact that human faces are highly, structurally symmetrical, learning-based hallucination methods can generally achieve superior performances to reconstruction-based hallucination methods, especially when the magnification factor is large; say, 4-8 times [4]. In this paper, we focus on learning-based face-hallucination algorithm based on a single input face image.

For most of current learning-based face-hallucination methods, they can generally be divided into two steps: global face reconstruction and residual face compensation. The first step is to model the global face appearance using subspace methods. However, the global faces obtained in this step always look blurred and lack facial features. Therefore, the second step is to compensate the reconstruction errors introduced in the first step, and to reduce blocky artifacts in the reconstructed images. Liu et al. [5] proposed a two-step statistical modeling approach, which combines a global parametric model and a local nonparametric model. This work has proven the necessity and efficiency of the two-step face-hallucination approach, and also proven that better global face will help achieve more reliable hallucinated faces.

In the global-face reconstruction phase, a subspace representation based on PCA [5], Kernel PCA [6], locality-preserving projection (LPP) [7], etc. is usually applied. However, directly applying these subspace transformations will lead to substantial differences between the projected LR and HR coefficients [8]. So, how to increase the correlation of the LR and HR manifolds remains to be studied. Among all the different methods, face hallucination based on canonical correlation analysis (CCA) has received promising results [8][9]. In the global reconstruction step, CCA is applied to project the PCA coefficients of both the LR and the HR training images into a coherent subspace, where the correlation between them is maximized. When given a novel LR face input, the corresponding reconstruction coefficients and weights are computed in the subspace to form the global HR face. However, regardless of whether 1D or 2D CCA is used in these works, the direction vectors learned are not orthogonal; this makes the reconstruction with pseudo-inverse lose some correlation, and thus makes the final global reconstruction less precise. We call this kind of CCA the 'original CCA' in this paper. To improve the performance, some constraints are added to the original CCA so as to obtain orthonormal direction matrices for perfect data reconstruction [10], and results have shown its superior performance on feature fusion.

Our proposed method is inspired by the previous CCA-based face-hallucination methods and the orthogonal variant of the original CCA method. We will show that, after including the orthogonality constraint on the original CCA, the direction vectors derived will be more correlated and will also facilitate reconstruction. Compared to [8], which simply maximizes the correlation between the HR and LR coefficients, our method can further improve the quality of the global reconstruction result. We believe that by obtaining better global face will further assist the overall face hallucination performance which holds a potential for real-life applications.

The remainder of this paper is organized as follows. In Section II, we briefly review the theory of CCA and the orthogonal CCA. We will present the realization of our orthogonal CCA-based global face reconstruction method in Section III. The experiment results and analysis are given in Section IV, and the conclusions are provided in Section V.

II. Related Work

In this section, we will give a brief overview of the CCA theory and its applications. We will also introduce the concept of orthogonal CCA as one of its important variants to show its efficiency for data reconstruction.

A. *Canonical Correlation Analysis (CCA)*

CCA is a learning method which seeks basis vectors for two sets of variables, say $\boldsymbol{x}$ and $\boldsymbol{y}$, such that their projections onto the basis vectors have a maximized correlation [11]. Denote $\mathbf{X}$ and $\mathbf{Y}$ as the matrices whose columns are the sets of variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with zero mean, respectively. Suppose that $\mathbf{M}$ and $\mathbf{N}$ are the respective direction matrices for $\mathbf{X}$ and $\mathbf{Y}$, and the corresponding canonical variate matrices of the projection coefficients are denoted as $\mathbf{U}$ and $\mathbf{V}$, i.e. $\mathbf{U}=\mathbf{M}^T\cdot\mathbf{X}$ and $\mathbf{V}=\mathbf{N}^T\cdot\mathbf{Y}$. Then, CCA maximizes the following correlation:

$$\rho=\frac{E[\mathbf{UV}]}{\sqrt{E[\mathbf{U}^2]\mathrm{E}[\mathbf{V}^2]}}=\frac{\mathbf{M}^T\mathbf{C}_{XY}\mathbf{N}}{\sqrt{\mathbf{M}^T\mathbf{C}_{XX}\mathbf{M}\cdot\mathbf{N}^T\mathbf{C}_{YY}\mathbf{N}}}, \tag{1}$$

where T represents the transpose operation; $\mathbf{C}_{XX}$ and $\mathbf{C}_{YY}$ denote the within-set covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively; and $\mathbf{C}_{XY}$ denotes the covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$. It can be shown that the optimal direction matrices $\mathbf{M}$ and $\mathbf{N}$ are the eigenvectors of $\mathbf{R}_1=\mathbf{C}_{XX}^{-1}\mathbf{C}_{XY}\mathbf{C}_{YY}^{-1}\mathbf{C}_{YX}$ and $\mathbf{R}_2=\mathbf{C}_{YY}^{-1}\mathbf{C}_{YX}\mathbf{C}_{XX}^{-1}\mathbf{C}_{XY}$, respectively.

B. *Orthogonal CCA*

In [10], the original CCA is extended to orthogonal CCA (oCCA). The orthogonality property is crucial for data reconstruction, and can make the PCA projections more consistent. Therefore, in our proposed face super-resolution framework, we also apply oCCA to impose extra constraints on the original CCA. The orthogonal direction matrices $\mathbf{M}$ and $\mathbf{N}$ can be computed in an iterative way as follows:

$$\arg\max_{\boldsymbol{m}_k,\boldsymbol{n}_k} \boldsymbol{m}_k^T\mathbf{C}_{XY}\boldsymbol{n}_k$$

$$\text{subject to}\begin{cases}\boldsymbol{m}_1^T\boldsymbol{m}_k=\boldsymbol{m}_2^T\boldsymbol{m}_k=\cdots=\boldsymbol{m}_{k-1}^T\boldsymbol{m}_k=0,\\ \boldsymbol{n}_1^T\boldsymbol{n}_k=\boldsymbol{n}_2^T\boldsymbol{n}_k=\cdots=\boldsymbol{n}_{k-1}^T\boldsymbol{n}_k=0,\\ \boldsymbol{m}_k^T\mathbf{C}_{XX}\boldsymbol{m}_k=1,\\ \boldsymbol{n}_k^T\mathbf{C}_{YY}\boldsymbol{n}_k=1,\end{cases} \tag{2}$$

where $\boldsymbol{m}_k$ and $\boldsymbol{n}_k$ are the kth column vectors of the direction matrices $\mathbf{M}$ and $\mathbf{N}$, respectively. The first two constraints are to impose orthogonality on the direction vectors, while the last two are additional constraints on the norm of $\boldsymbol{m}_k$ and $\boldsymbol{n}_k$. The details of deriving the direction vectors can be found in [10]. Having computed the orthogonal-direction matrices, we can further normalize them to become orthonormal.

To visualize the correlation between the projected HR and LR coefficients of different subspace methods, we show their lower-dimensional geometry using LLE [12] from 800 randomly chosen training samples. As illustrated in Fig. 1, the difference between the HR and LR manifolds of the PCA coefficients is the largest. Both the original CCA and the orthogonal CCA can preserve the intrinsic correlation well based on the similarity between the LR and HR manifolds. This shows that using orthogonal transformation will not reduce the correlation, and it can also provide a better performance in terms of reconstruction, which will be discussed later.

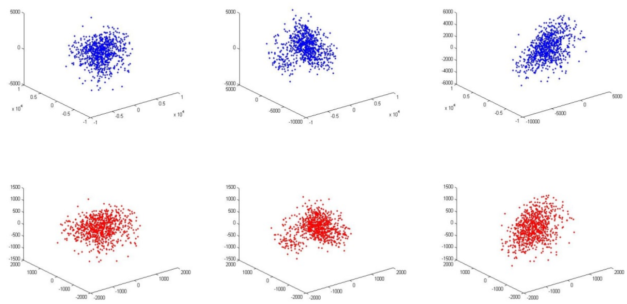

(a) (b) (c)

Fig. 1. Comparison of the correlation between the manifolds of the projected LR and HR coefficients based on (a) PCA, (b) original CCA, and (c) proposed orthogonal CCA. The first row shows the manifolds of the projected LR coefficients, while the second row shows those of the projected HR coefficients.

III. Global Face Reconstruction Based on Orthogonal CCA

In this section, we will present our global face reconstruction framework based on the proposed orthogonal CCA. PCA is first applied to both the LR and HR demeaned training faces. This is necessary because some noise can be removed, and the dimensionality of the face samples can be reduced so that the matrices involved in the orthogonal CCA are non-singular and the direction matrices can be computed more efficiently.

Suppose that the LR and HR training face images are represented as $\mathbf{I}_X$ and $\mathbf{I}_Y$, respectively, which are in the face form of matrices, with each column representing one face. The corresponding mean faces of the LR and HR training images are denoted as μ_X and μ_Y , respectively. By applying PCA, the orthonormal eigenvector matrices $\mathbf{E}_X$ and $\mathbf{E}_Y$ can be obtained, with the leading eigenvectors containing 98% of the total variation. The projection coefficients $\mathbf{B}_X$ and $\mathbf{B}_Y$ can be computed as follows:

$$\mathbf{B}_X = \mathbf{E}_X^T(\mathbf{I}_X - \mu_X), \text{ and}$$

$$\mathbf{B}_Y = \mathbf{E}_Y^T(\mathbf{I}_Y - \mu_Y). \quad (3)$$

It can be proven that the two coefficient matrices are also zero-centered. Thus, we can directly apply the orthogonal CCA to $\mathbf{B}_X$ and $\mathbf{B}_Y$, as described in Section II. After rescaling, two orthonormal direction matrices $\mathbf{P}_X$ and $\mathbf{P}_Y$, as well as two projected coefficient matrices $\mathbf{O}_X$ and $\mathbf{O}_Y$ with increased correlation, can be obtained, i.e.

$$\begin{aligned}\mathbf{O}_X &= \mathbf{P}_X^T \cdot \mathbf{B}_X = \mathbf{P}_X^T \cdot \mathbf{E}_X^T(\mathbf{I}_X - \mu_X) \\ &= (\mathbf{E}_X\mathbf{P}_X)^T(\mathbf{I}_X - \mu_X)\end{aligned} \quad (4)$$

$$\begin{aligned}\mathbf{O}_Y &= \mathbf{P}_Y^T \cdot \mathbf{B}_Y = \mathbf{P}_Y^T \cdot \mathbf{E}_Y^T(\mathbf{I}_Y - \mu_Y) \\ &= (\mathbf{E}_Y\mathbf{P}_Y)^T(\mathbf{I}_Y - \mu_Y)\end{aligned} \quad (5)$$

From these equations, we can see that, after applying orthogonal CCA to the PCA coefficients, we actually project the input samples on to an orthonormal, consistent subspace, which is formed by multiplying the orthonormal projection matrices on the right-hand side of original eigenvectors. We denote $\tilde{\mathbf{E}}_X = \mathbf{E}_X\mathbf{P}_X$ and $\tilde{\mathbf{E}}_Y = \mathbf{E}_Y\mathbf{P}_Y$. Then, we can verify that

$$\tilde{\mathbf{E}}_X^T \cdot \tilde{\mathbf{E}}_X = (\mathbf{E}_X\mathbf{P}_X)^T \cdot (\mathbf{E}_X\mathbf{P}_X) = \mathbf{I}, \text{ and}$$

$$\tilde{\mathbf{E}}_Y^T \cdot \tilde{\mathbf{E}}_Y = (\mathbf{E}_Y\mathbf{P}_Y)^T \cdot (\mathbf{E}_Y\mathbf{P}_Y) = \mathbf{I}, \quad (6)$$

where $\mathbf{I}$ is an identity matrix. The projected eigenvector matrices remain orthonormal, so they can be used directly for reconstruction. None of the existing face-hallucination methods based on the original CCA possesses this feature.

To reconstruct the global face, we compute the PCA coefficients $\boldsymbol{b}_l$ of the input LR face image $\boldsymbol{I}_l$. Then, these coefficients are projected on to a more correlated subspace using orthogonal CCA:

$$\boldsymbol{o}_l = \mathbf{P}_X^T \cdot \boldsymbol{b}_l. \quad (7)$$

Since human faces have a similar structure and texture, the projected coefficients of the input LR face can be represented as a linear combination of the projected coefficients of its K nearest neighbors $\{\boldsymbol{o}_{Xj}\}_{j=1}^K$ from the training LR face images $\mathbf{O}_X$. This is realized by minimizing the reconstruction errors as follows:

$$\varepsilon = \left|\boldsymbol{o}_l - \sum_{j=1}^K w_j^g \boldsymbol{o}_{Xj}\right|^2, \text{ subject to } \sum_{j=1}^K w_j^g = 1, \quad (8)$$

where $\boldsymbol{o}_{Xj}$ is the jth nearest neighbor of the input LR face, and w_j^g is the corresponding weight for $\boldsymbol{o}_{Xj}$. This is a constrained least-squares fit problem, which can be solved by using the method in [13].

As mentioned previously, the basic assumption of the learning-based face-hallucination approach is that the same neighborhoods are preserved in both the HR and the LR manifolds. Based on this assumption, the corresponding projected PCA coefficients of a desired global HR face can be represented as a linear combination of its K nearest neighbors $\{\boldsymbol{o}_{Yj}\}_{j=1}^K$ in $\mathbf{O}_Y$, using the same weight contributions as follows:

$$\boldsymbol{o}_h = \sum_{j=1}^K w_j^g \boldsymbol{o}_{Yj}. \quad (9)$$

Similar to (4) and (5), $\boldsymbol{o}_h$ is also related to the global HR face as follows:

$$\boldsymbol{o}_h = \mathbf{P}_Y^T \cdot \boldsymbol{b}_h = \mathbf{P}_Y^T \cdot \mathbf{E}_Y^T(\boldsymbol{I}_h^g - \boldsymbol{\mu}_Y) = \tilde{\mathbf{E}}_Y^T(\boldsymbol{I}_h^g - \boldsymbol{\mu}_Y), \quad (10)$$

where $\boldsymbol{b}_h$ represents the PCA coefficients of the global HR face computed by projecting its demeaned face onto the HR eigenvector matrix. Thus, the global HR face can be obtained as follows:

$$\boldsymbol{I}_h^g = \tilde{\mathbf{E}}_Y \cdot \boldsymbol{o}_h + \boldsymbol{\mu}_Y. \quad (11)$$

IV. EXPERIMENT AND ANALYSIS

To compare our proposed global reconstruction framework with other face-hallucination methods, the CAS-PEAL-R1 dataset [14] was used. This dataset contains 1,040 individuals with different poses, illumination, backgrounds, etc. In our experiments, we chose all 1,040 frontal faces with neutral expression and normal lighting from the dataset. All of these faces are cropped to include the face region only and are aligned based on the two eye centers to form the HR face images of size 128×128 pixels. Then, all the HR faces are blurred with low-pass Gaussian filtering and are down-sampled to the size of 32×32 pixels. For experiment evaluation, we randomly selected 1,000 face images as a training set and the remaining 40 face images as a testing set in each experiment. The experiment is repeated five times, and the average results are measured.

In this section, we will firstly compare the global face reconstruction results of our proposed method with other state-of-the-art face-hallucination methods. Then, we will discuss some of the important parameters used in our method.

A. Global face reconstruction

In order to evaluate our orthogonal CCA method in terms of the performance of global face reconstruction, we compare it with another three related methods: original CCA [8], 2D CCA [9], and LPH [7]. In [8], the original CCA algorithm is

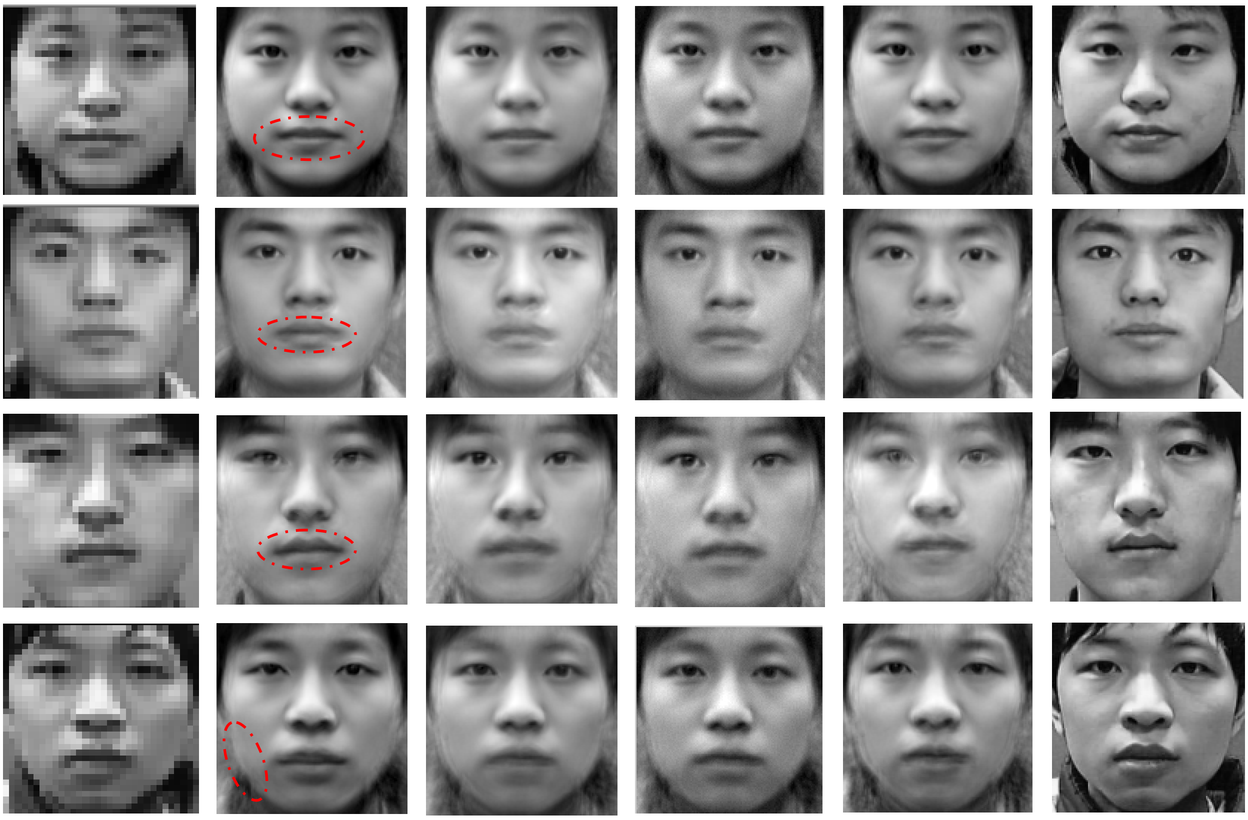


Fig. 2. Global face-reconstruction results: (a) input LR faces, (b) global faces produced by our proposed orthogonal CCA, (c) global faces reconstructed using original CCA in [8], (d) global faces reconstructed using 2D CCA in [9], (e) global faces reconstructed using LPH in [7], and (f) the original HR faces.

applied to increase the correlation of the HR and LR coefficients without considering the reconstruction issue. In [9], 2D CCA is directly applied to the training and testing image data without performing vectorization and PCA. In [7], global faces are reconstructed based on the LPP method, followed by the use of the radial-basis-function regression.

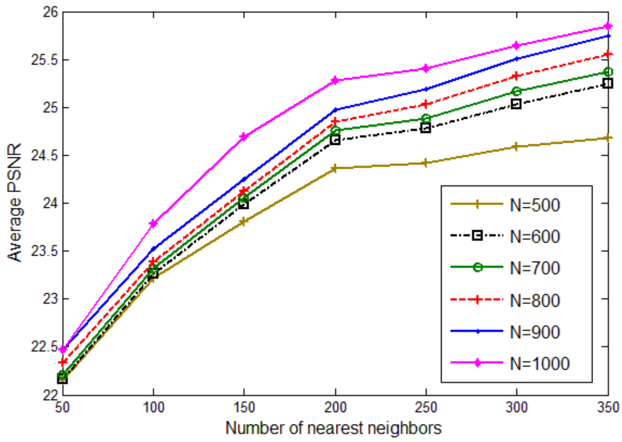
For the original CCA and LPH methods, we set all the parameters the same as in [9]. We also follow the same setting of 2D CCA in [9], which partitions faces into three parts and applies 2D CCA to each part for reconstruction. For our method, leading eigenvectors of 98% of the total variation are used, and the number of nearest neighbors K is set at 200 in the nearest-neighbor searching.

Figure 2 shows the global face-reconstruction results based on the different methods. It can be seen that our proposed method can achieve the best performance in terms of visual quality. It should be noted that we have aligned the face images roughly based on their eye positions only. However, this rough alignment leads to the results based on the original CCA, 2D CCA, and LPH suffering from a severe jagged effect around the mouth and the chin regions in the reconstructed face images. However, the global faces reconstructed using orthogonal CCA exhibit much less distortion in these regions and a better visual quality, as indicated by the red-dashed circles in Fig. 2(b).

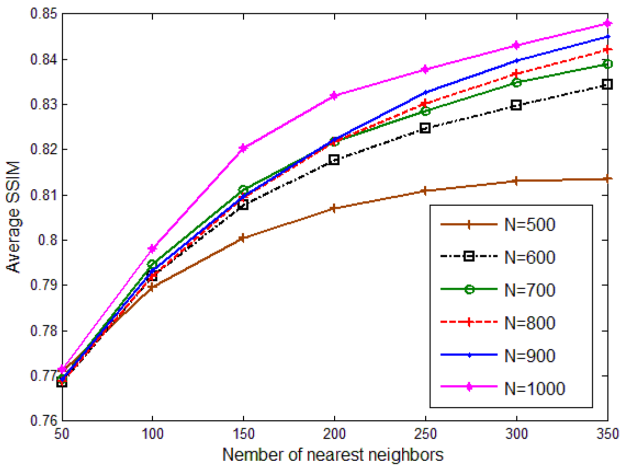
Table 1 Global-face reconstruction performances in terms of the means and standard deviations (mean±standard deviation) of the PSNR and SSIM of the different methods.

	Orthogonal CCA	Original CCA	2D CCA	LPH
PSNR	**25.57 ± 1.08**	23.05 ± 2.66	23.66 ± 2.64	23.87 ± 2.61
SSIM	**0.831 ± 0.02**	0.792 ± 0.04	0.812 ± 0.04	0.804 ± 0.04

We also compare our proposed orthogonal CCA with the original CCA and LPH for global-face reconstruction in terms of the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [15]. In total, we produced 200 reconstructed global faces after five trials. Then, the averages and standard deviations of the PSNR and SSIM of the three methods are measured, and the results are tabulated in Table 1. From the results, we can see an obvious improvement in both measurements with our method. Moreover, the results show that the orthogonal CCA method can reconstruct global faces with a more stable and uniform performance, as the standard deviations based on our method are smaller than other methods; this further shows the advantage of including the orthogonality property in reconstruction. Global faces reconstructed using 2D CCA achieve better visual results, but the average PSNR is lower because applying projection directly onto face images inevitably introduces noise.



(a)



(b)

Fig. 3. (a) The average PSNRs, and (b) average SSIMs of the reconstructed global faces based on our method under different training-set sizes and numbers of nearest neighbors.

B. Parameters for global face reconstruction

For our proposed global face-reconstruction method, we investigate the effects of the training-set size and the number of nearest neighbors searched on the reconstruction performance. The size of the training set N is changed from 500 to 1,000, with an interval of 100. For each of the training-set sizes, the number of nearest neighbors K varies from 50 to 350, with an interval of 50. The average PSNR and SSIM values of the 40 test face images for each case are computed and displayed in Fig. 3. It can be seen that, at each training size, increasing the number of nearest neighbors will lead to a higher PSNR and SSIM. However, if we search for a larger number of nearest neighbors, the computational cost will increase. From the experiment results, there will be a slight improvement in performance when K is larger than 200. Therefore, in our algorithm, K is set at 200. On the other hand, increasing the training-set size will constantly improve the reconstruction performance of our method. In the training phase, learning the orthogonal CCA direction matrices iteratively requires most of the computation. However, once the matrices have been learned, they can be used directly for the reconstruction of novel LR face images. More training samples can help the learning with orthogonal CCA becoming more accurate. As a result, we keep using the largest training set, i.e. N=1,000, in our method. Figure 4 shows some reconstructed global faces, with K=200 and different training-set sizes.

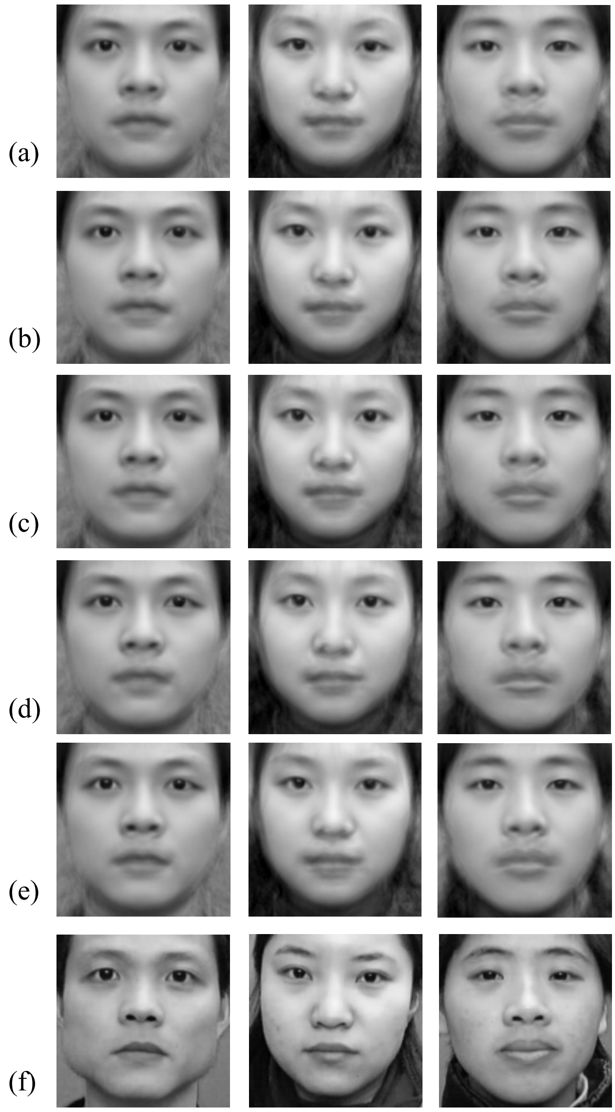


Fig. 4. Globally reconstructed faces with different training-set sizes N, and K=200: (a) N=600, (b) N=700, (c) N=800, (d) N=900, (e) N=1,000, and (f) the original HR faces.

V. CONCLUSIONS

In this paper, a global face reconstruction framework for face hallucination based on orthogonal CCA has been proposed. By applying orthogonal transformation to the original CCA, the orthogonal CCA proposed becomes more efficient for data reconstruction. Experiments have shown a great improvement in global face reconstruction using our method. Our future work will focus on studying an efficient compensation framework to produce hallucinated faces with higher visual quality and applying the method to real-life applications.

REFERENCES

[1] S. Baker, T. Kanade, Hallucinating faces, in: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 83–88.

[2] M. Elad and A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, IEEE Transactions on Image Processing, 1997, 6(12): 1646–1658.

[3] H. He, L. P. Kondi, An image super-resolution algorithm for different error levels per frame, IEEE Transactions on Image Processing, 2006, 15(3): 592-603.

[4] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(9): 1167-1183.

[5] C. Liu, H. Y. Shum, W. T. Freeman, Face hallucination: Theory and practice, International Journal of Computer Vision, 2007, 75(1): 115-134.

[6] Y. Hu, K. M. Lam, T. Shen, W. Wang, A novel kernel-based framework for facial-image hallucination, Image and Vision Computing, 2011, 29(4): 219-229.

[7] Y. Zhuang, J. Zhang, F. Wu, Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation, Pattern Recognition, 2007, 40(11): 3178-3194.

[8] H. Huang, H. He, X. Fan, J. Zhang, Super-resolution of human face image using canonical correlation analysis, Pattern Recognition, 2010, 43(7): 2532-2543.

[9] L. An, B. Bhanu, Face image super-resolution using 2D CCA, Signal Processing, 2013.

[10] X. B. Shen, Q. S. Sun, Y. H. Yuan, Orthogonal canonical correlation analysis and its application in feature fusion, IEEE International Conference on Information Fusion (FUSION), 2013: 151-157.

[11] S. Szedmak, J. T. Shawe, D. Hardoon, Canonical correlation analysis: An overview with application to learning methods, Neural computation, vol. 16, no. 12, pp. 2639-2664, 2004.

[12] S. T. Roweis, L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science, 2000, 290(5500): 2323-2326.

[13] L. K. Saul, S. T. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, The Journal of Machine Learning Research, 2003, 4: 119-155.

[14] W. Gao, B. Cao, S. Shan, The CAS-PEAL large-scale Chinese face database and baseline evaluations, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 2008, 38(1): 149-161.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing, 2004, 13(4): 600-612.