# Face Hallucination Through Dual Associative Learning

Wei Liu, Dahua Lin, and Xiaoou Tang
Department of Information Engineering
The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong SAR
Email: {wliu, dhlin4, xtang}@ie.cuhk.edu.hk

*Abstract*— **In this paper, we propose a novel patch-based face hallucination framework, which employs a dual model to hallucinate different components associated with one facial image. Our model is based on a statistical learning approach: *Associative Learning*. It suffices to learn the dependencies between low-resolution image patches and their high-resolution ones with a new concept *Hidden Parameter Space* as a bridge to connect those patches with different resolutions. To compensate higher frequency information of images, we present a dual associative learning algorithm for orderly inferring main components and high frequency components of faces. The patches can be finally integrated to form a whole high-resolution image. Experiments demonstrate that our approach does render high quality super-resolution faces.**

## I. Introduction

In face identification and recognition applications, it is generally useful to render a high-resolution face image from the low-resolution one, which is one popular topic in computer vision community and called face hallucination or face super-resolution. A great number of super-resolution techniques have been proposed in recent years [1][2][3][5][6][7][8][10]. Most target at producing a super-resolution image from a sequence of low-resolution images [5][7][10]. Some other approaches are based on learning from the training set containing high- and low-resolution image pairs, with the assumption that high-resolution images are Markov Random Field (MRF) [3][6][7]. These methods are more suitable for synthesizing local texture, and are usually applied to generic images without special consideration on the property of face images.

Baker and Kanade [1][2] develop a hallucination method based on the property of face image. Abandoning the MRF assumption, it infers the high frequency components from a parent structure by recognizing the local features from the training set. Liu *et al.*[8] develop a two-step statistical approach integrating global and local models. Both of the two methods use complicated probabilistic models and are based on an explicit resolution reduction function, which is sometimes difficult to obtain in practice.

Instead of using a probabilistic model, Wang *et al.* [14] propose a face hallucination method directly based on linear transform between two spaces of different dimensions. Though simple, as a global linear approach, it suffers the serious problem of losing important detail information.

In this paper, we propose a novel framework based on image patches for solving single facial image super-resolution problems. We develop the dual associative learning model to hallucinate different components associated with one facial image.

## II. Related Work

Given the high dimensional nature of images, modeling non-linear dependencies in the image space is often infeasible due either to limited training data or computational complexity. While linear relationships beyond PCA are hard to model due to high dimensional data with the lack of training samples. That presents a number of challenges.

Suppose that two related data sets $\mathbf{D} \in \Re^{d_1 \times n}$ and $\hat{\mathbf{D}} \in \Re^{d_2 \times n}$ with an equal number of observations are available, it arises that how to exploit the correlations between them to estimate one set from the other. This problem has been studied in the signal processing community [12] and neural network community [4]. It is known as reduced rank Wiener filtering [12] or Asymmetric PCA (APCA) [4]. In contrast to joint PCA, the purpose of Asymmetric PCA is to explore the use of linear models for learning relations between two given data sets while coupling the coefficients in a robust way. For simplicity, we assume that both $\mathbf{D}$ and $\hat{\mathbf{D}}$ are zeros mean. APCA can be formulated as the minimization of

$$E_{apca}(\mathbf{B}, \hat{\mathbf{B}}) = \sum_{i=1}^{n} \|\hat{\mathbf{d}}_i - \hat{\mathbf{B}}\mathbf{B}^{\mathrm{T}}\mathbf{d}_i\|^2 \quad (1)$$

From above equation, we can observe that if $\hat{\mathbf{d}}_i = \mathbf{d}_i$ and $\hat{\mathbf{B}} = \mathbf{B}$ then minimizing (1) leads to the standard (symmetric) PCA [4]. When working with high dimensional data such as images, the solution of (1) is hard to achieve.

Torre *et al.* [13] extend APCA and formulate Asymmetric Coupled Component Analysis (ACCA) in such a way that the hidden coefficients are made explicit. This differs from and generalizes previous work in that it allows us to impose constraints on the coupling. Although the solution of ACCA can be derived from a more robust objective function in [13], the basis philosophy of ACCA is not applicable to super-resolution problems. The reason is that there exists a formidable challenge in that applying ACCA to learn the relationships between low- and high-resolution image data sets with large dimension gap. ACCA might fail in the particular case of high difference in dimensions of two data sets.

## III. Associative Learning

The purpose of this paper is to describe a associative learning method for learning dependencies between low-resolution image data sets and high-resolution data sets in the hidden parameter space rather than the observation space. The proposed associative learning method is directly applied to hallucinating faces. To assure hallucinating quality, we develop a dual associative learning framework to reinforce once associative learning performance.

### A. Patch-Based Asymmetric Associative Learning

The fundamental task of face hallucination is to learn the relationship between low-resolution images and high-resolution ones. However, the dimension of image space is too high to effectively establish this relation. To address the problem, we divide the whole face image into a set of small overlapped small patches as elements for hallucinating. In order to establish the relationship between the low-resolution patch space and the high resolution patch space, we develop a statistical learning method named *Patch-Based Asymmetric Associative Learning*, which associates and correlates two patch spaces by introducing an intermediate space.

Denote the low resolution patch space as $\Omega_L \in \Re^{d_L}$ and the high resolution patch space as $\Omega_H \in \Re^{d_H}(d_L < d_H)$. We use $\mathbf{x}_L$ to denote the vector in space $\Omega_L$, and use $\mathbf{x}_H$ to denote the vector in space $\Omega_H$. Suppose we have $n(>> d_H > d_L)$ training samples, then we can arrange all these training samples into two matrices: $\mathbf{X}_L = [\mathbf{x}_L^{(1)}, \cdots, \mathbf{x}_L^{(n)}]$ and $\mathbf{X}_H = [\mathbf{x}_H^{(1)}, \cdots, \mathbf{x}_H^{(n)}]$, where $\mathbf{x}_L^{(i)}$ is the vector representation of $i$-th low resolution patch, while $\mathbf{x}_H^{(n)}$ is the vector representing the corresponding high resolution part. For succinct formulation, we enforce $\mathbf{X}_L$ and $\mathbf{X}_H$ to have zero means by subtracting the mean vector respectively.

Assuming there exists a linear relation between the two spaces, we can approximate $\mathbf{x}_H$ by $\mathbf{A}\mathbf{x}_L$. The problem of solving $\mathbf{A}$ can be formulated as a multivariate regression problem as

$$E_{reg}(\mathbf{A}) = \sum_{i=1}^{n} \|\mathbf{x}_H^{(i)} - \mathbf{A}\mathbf{x}_L^{(i)}\|^2, \quad (2)$$

because $\mathbf{X}_L\mathbf{X}_L^{\mathrm{T}}$ is invertible, the closed form solution to above problem is $\mathbf{A} = \mathbf{X}_H\mathbf{X}_L^{\mathrm{T}}(\mathbf{X}_L\mathbf{X}_L^{\mathrm{T}})^{-1}$.

Nonetheless, in real application, it is always the case that samples in $\Omega_L$ or $\Omega_H$ only distribute in a subspace, which allows us to impose rank constraint on the matrix $\mathbf{A}$ in regression. Intuitively speaking, the rank of $\mathbf{A}$ is an important parameter characterizing the complexity of the linear model. To strike a well balance between fidelity and generalization ability and robustly estimate the relationship, it is crucial to identify the subspace. Based on this observation, we integrate the multivariate linear regression and the well-known dimensionality reduction method SVD (PCA) by introducing the concept *Hidden Parameter Space*.

In our formulation, the vector $\mathbf{x}_L^{(i)}$ in space $\Omega_L$ and the vector $\mathbf{x}_H^{(i)}$ in space $\Omega_H$ correspond to the same vector $\mathbf{h}_i$ in hidden parameter space. In ideal case without noise, we have

$$\mathbf{x}_L^{(i)} = \mathbf{B}_L\mathbf{h}_i, \qquad \mathbf{B}_L \in \Re^{d_L \times k} \quad (3)$$

$$\mathbf{x}_H^{(i)} = \mathbf{B}_H\mathbf{h}_i, \qquad \mathbf{B}_H \in \Re^{d_H \times k} \quad (4)$$

The dimension of hidden parameter space is smaller than that of space $\Omega_L$ and space $\Omega_H$: $k < d_L < d_H$. In addition, the base matrices $\mathbf{B}_L$ and $\mathbf{B}_H$ are orthogonal matrices. Then the linear relation between two spaces is as below,

$$\mathbf{x}_H = \mathbf{B}_H\mathbf{B}_L^{\mathrm{T}}\mathbf{x}_L. \quad (5)$$

Combining (5) and (2), the associative learning model can be finally formulated as minimization of the following energy function

$$E(\mathbf{B}_L, \mathbf{B}_H) = \sum_{i=1}^{n} \|\mathbf{x}_H^{(i)} - \mathbf{B}_H\mathbf{B}_L^{\mathrm{T}}\mathbf{x}_L^{(i)}\|^2. \quad (6)$$

By representing $\mathbf{A}$ by $\mathbf{B}_H\mathbf{B}_L^{\mathrm{T}}$, we indeed impose an rank constraint on $A$: $\mathrm{rank}(\mathbf{A}) \leq k$. According to the theory of Singular Value Decomposition, $\mathbf{B}_H$ and $\mathbf{B}_L$ can be solved by performing SVD on $\mathbf{A}$(e.g. $\mathbf{A} = \mathbf{U\Sigma V}^{\mathrm{T}}$), and computing $\mathbf{B}_L$ and $\mathbf{B}_H$ by selecting the first $K(\leq k)$ left singular vectors $\mathbf{U}$ and $K$ right singular vectors $\mathbf{V}$ and scaling them by square root of corresponding singular values; that is

$$\begin{aligned} \mathbf{B}_H &= \dot{\mathbf{U}}\dot{\mathbf{\Sigma}}^{1/2} \\ \mathbf{B}_L &= \dot{\mathbf{V}}\dot{\mathbf{\Sigma}}^{1/2} \end{aligned} \quad (7)$$

where $\dot{\mathbf{\Sigma}}$ is the top-left $K \times K$ submatrix of the pseudo diagonal matrix $\mathbf{\Sigma}$, and $\dot{\mathbf{U}}$ and $\dot{\mathbf{V}}$ both contain first $K$ columns of orthogonal matrices $\mathbf{U}$, $\mathbf{V}$.

### B. Dual Associative Learning

We develop the dual associative learning model to hallucinate different components associated with one facial image. The first learning model can recover global face structure and main local facial features. The second one accounts for learning higher frequency components (e.g. corners of eyes, mouth, nose and other facial details) of faces. Both models are learned from existing high-resolution face images and their smoothed and down-sampled lower resolution ones.

First, given the input training data sets $\{I_L^{(i)}\}_{i=1}^{l}$(low-resolution face data set) and $\{I_H^{(i)}\}_{i=1}^{l}$(high-resolution face data set), where $l$ is input sample number, we propose the *Associative Learning* algorithm:

- **Step 1**. Calculate pairwise mean face $\mu_L$ and $\mu_H$ from the training data sets.
- **Step 2**. Divide whole images $\{I_L^{(i)}\}_{i=1}^{l}$, $\{I_H^{(i)}\}_{i=1}^{l}$, $\mu_L$ and $\mu_H$ into a set of small overlapping image patch sets $\{\mathbf{p}_{L,j}^{(i)} | \mathbf{p}_{L,j}^{(i)} \in I_L^{(i)}, 1 \leq j \leq m, 1 \leq i \leq l\}$, $\{\mathbf{p}_{H,j}^{(i)} | \mathbf{p}_{H,j}^{(i)} \in I_H^{(i)}, 1 \leq j \leq m, 1 \leq i \leq l\}$, $\{\mu_{L,j}\}_{j=1}^{m}$ and $\{\mu_{H,j}\}_{j=1}^{m}$. Adopt same number and relative positions between low-resolution patches and their counterpart high-resolution

ones, $m$ is the share number of patches belong to one face image.

- **Step 3**. For each position $j = 1, \cdots, m$ of patches: Construct pairwise patch sets $\mathbf{X}_{L,j} = \{\mathbf{p}_{L,j}^{(i)} - \mu_{L,j}\}_{i=1}^{l}$ and $\mathbf{X}_{H,j} = \{\mathbf{p}_{H,j}^{(i)} - \mu_{H,j}\}_{i=1}^{l}$, and perform patch-based asymmetric associative learning to acquire the linear relation (reflected in matrices $\mathbf{B}_{L,j}$ and $\mathbf{B}_{H,j}$) between $\mathbf{X}_{L,j}$ and $\mathbf{X}_{H,j}$. Keep the final linear transformation matrix as $\mathbf{T}_j = \mathbf{B}_{H,j}\mathbf{B}_{L,j}^{\mathrm{T}}$

The associative learning algorithm outputs useful data $\{\mu_{L,j}\}_{j=1}^{m}$, $\{\mu_{H,j}\}_{j=1}^{m}$ and $\{\mathbf{T}_j\}_{j=1}^{m}$ for hallucinating faces.

For testing stage, we also divide the testing image $I_L$ into overlapping patches $\{\mathbf{p}_{L,j}\}_{j=1}^{m}$ as in training stage. Apply (5) to render the hallucinated high-resolution patch $\mathbf{p}_{H,j}$ for each patch $\mathbf{p}_{L,j}$. Concatenate and integrate the hallucinated high-resolution patches to form one facial image, which is just the target high-resolution facial image, with local compatibility and smoothness constraints. That is

$$\tilde{I}_H = \bigcup_{j=1}^{m} [\mathbf{T}_j(\mathbf{p}_{L,j} - \mu_{L,j}) + \mu_{H,j}] \tag{8}$$

where the operator $\bigcup$ represents concatenating and integrating the patches belong to one face image to form a full face image, simultaneously blending pixels in the overlapping area.

Generally, only single associative learning is not enough to hallucinate satisfactory results. Second associative learning is applied again to establish the relation between low resolution image residual and high resolution image residual caused by the first associative learning. Once the model is trained, we can infer the high-resolution residue from low-resolution residue, and further enhance the quality of hallucination. We combine the two-stage associative learning as a integrated framework, called *Dual Associative Learning*:

1) Split the training sets $\{I_L^{(i)}\}_{i=1}^{n}$ and $\{I_H^{(i)}\}_{i=1}^{n}$ into two disjoint halves. Send the first half to the associative leaning model, $\{\mu_{L,j}\}_{j=1}^{m}$, $\{\mu_{H,j}\}_{j=1}^{m}$ and $\{\mathbf{T}_j^{g}\}_{j=1}^{m}$ are acquired.
2) Using the low-resolution images of the remaining half training data as test images, hallucinate faces directly using (8) by exploiting the results from step 1. Then two residual images for each sample can be constructed as below: one is obtained by subtracting the low-resolution image by a down-sampled version of the hallucinated image, the other is obtained by subtracting the actual high-resolution image by the hallucinated image.
3) Input low- and high-resolution residue images to the associative leaning model to compute $\{\nu_{L,j}\}_{j=1}^{m}$, $\{\nu_{H,j}\}_{j=1}^{m}$ and $\{\mathbf{T}_j^{r}\}_{j=1}^{m}$.

For an input low-resolution image $I_L$ with its overlapping patches $\{\mathbf{p}_{L,j}\}_{j=1}^{m}$, the global super-resolution image $I_H^{g}$ is hallucinated from (8) by substituting the formula with $\mathbf{p}_{L,j}$, $\mu_{L,j}$, $\mu_{H,j}$ and $\mathbf{T}_j^{g}$. Construct low-resolution residue image $I_L^{r}$ with its overlapping patches $\{\mathbf{q}_{L,j}\}_{j=1}^{m}$, by subtracting the input image $I_L$ with the down-sampled version of hallucinated image $I_H^{g}$. As a same way for hallucinating $I_H^{g}$, substitute (8) with $\mathbf{q}_{L,j}$, $\nu_{L,j}$, $\nu_{H,j}$ and $\mathbf{T}_j^{r}$ to infer the high-resolution residue image $I_H^{r}$. Add the inferred residue image $I_H^{r}$ to the global version $I_H^{g}$ to render the final result $I_H^{*} = I_H^{g} + I_H^{r}$.

## IV. Experiments

Our experiments were conducted using a mixed database, which is a collection of two databases **XM2VTS** [9] and **FERET** [11]. Our training data set consists of about 1400 images. Among all these samples, we select a half part samples for training the global model and the remain part for training residue compensation. Other samples and some outside samples are for testing. As a necessary preamble steps, we perform geometric normalization by an affine transform based on coordinates of eyes and mouth. After the transform, each image is cropped to a canonical $96 \times 128$ grayscale image as the high-resolution one. The corresponding low-resolution images are obtained by smoothing and down-sampling, which are $24 \times 32$ images.

In our experiments, for each low-resolution image, 682 (that is the value for $m$) overlapped patches are extracted by sliding a small $3 \times 3$ window pixel by pixel. For high-resolution images, 682 $12 \times 12$ patches are extracted as well. The patches in low- and high-resolution image are in one-to-one correspondence. The dimension $k$ of the hidden parameter space is set to 5, the parameter $K$ is also set to 5. Our experiments show that such configuration on parameters yields the most satisfactory results. Notice that the two parts for training the dual associative learning model must be disjoint from each other.

The resultant images are shown in Fig. 1. We can see that single associative learning (plotted in Fig. 1(d)) can produce good hallucinated results, and dual associative learning (plotted in Fig. 1(e)) further enhances the quality. The hallucinated image approximates the groundtruth fairly well.

## V. Conclusion

We have compared our algorithm with other existing methods, including Cubic B-Spline and Baker's algorithms, the results are shown in Fig. 1 from which we can clearly see the limitation of other methods. The quality of cubic B-Spline reconstruction is rather rough. Baker et al's method produces a better but still fairly blurred face. Our method has advantages over others in terms of preserving both global structure and subtle details.

## Acknowledgment

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region. The work was done while all the authors are with the Chinese University of Hong Kong. The authors thank Qingshan Liu for his constructive suggestions. And great appreciations to Yilin Dong for her encourage.

(a) Input 24*32 (b) Cubic B-Spline (c) Baker et al (d) Single Associative Learning (e) Dual Associative Learning (f) Original 96*128

Fig. 1. Comparison between our method and others.

## REFERENCES


[1] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break them," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 9, pp. 1167-1183, 2002.

[2] S. Baker and T. Kanade, "Hallucinating Faces," in *Proc. of Inter. Conf. on Automatic Face and Gesture Recognition*, pp. 83-88, 2000.

[3] J.D. Bonet, "Multiresolution Sampling Procedure for Analysis and Synthesis of Texture Images," *Proc. of SIGGRAPH 97*, pp. 361-368, 1997.

[4] K. Diamantaras, "Principal Component Neural Networks (Theory and Applications)," . *John Wiley & Sons*, 1996.

[5] M. Elad and A. Feuer, "Super-Resolution Reconstruction of Image Sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 9, 1999.

[6] W.T. Freeman and E.C. Pasztor, "Learning Low-Level Vision," in *Proc. of ICCV*, Vol. 2, pp. 1182-1189, 1999.

[7] R. Hardie, K. Barnard, and E. Armstrong, "Joint MAP registration and highresolution image estimation Using a sequence of undersampled images," *IEEE Trans. on Image Processing*, Vol. 6, No. 12, pp. 1621-1633, 1997.

[8] C. Liu, H. Shum, and C. Zhang, "A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model," in *Proc. of CVPR*, Vol. 1, pp. 192-198, 2001.

[9] K.Messer, J.Matas, J.Kittler, J.Luettin, and G.Matitre, "XM2VTSDB: The extended M2VTS database," in *Proc. of the Second International Conference on AVBPA*, 1999.

[10] A. Patti, M. Sezan, and A. Tekalp, "Super-resolution Video Reconstruction with Arbitrary Sampling Latices and Nonzero Aperture Time," *IEEE Trans. on Image Processing*, Vol. 6, No. 8, pp. 1064-1076, 1997.

[11] P. Philips, H. Moon, P. Pauss, and S.Rivzvi, "The FERET Evaluation Methodology for Face-Recognition Algorithms," in *Proc. of CVPR*, pp. 137-143, 1997.

[12] L. Scharf, "The SVD and reduced rank signal processing. SVD and Signal Processing, II," *(R. Vaccaro, ed.), Elsevier*, 1991.

[13] F.D. Torre and M.J. Black, "Dynamic Coupled Component Analysis," *Proc. of CVPR*, Vol. 2, pp. 643-650, 2001.

[14] X. Wang and X. Tang, "Hallucinating face by eigentransformation," to appear in *IEEE Trans. on Systems, Man, and Cybernetics*, Part-C, Special issue on Biometrics Systems, 2005.