Hallucinating Faces: TensorPatch Super-Resolution and Coupled Residue Compensation

Wei Liu[1], Dahua Lin[1], and Xiaoou Tang[1, 2]
[1]Department of Information Engineering
The Chinese University of Hong Kong, Shatin, Hong Kong
{wliu, dhlin4, xtang}@ie.cuhk.edu.hk
[2]Microsoft Research Asia, Beijing, China

Abstract

In this paper, we propose a new face hallucination framework based on image patches, which integrates two novel statistical super-resolution models. Considering that image patches reflect the combined effect of personal characteristics and patch-location, we first formulate a TensorPatch model based on multilinear analysis to explicitly model the interaction between multiple constituent factors. Motivated by Locally Linear Embedding, we develop an enhanced multilinear patch hallucination algorithm, which efficiently exploits the local distribution structure in the sample space. To better preserve face subtle details, we derive the Coupled PCA algorithm to learn the relation between high-resolution residue and low-resolution residue, which is utilized for compensate the error residue in hallucinated images. Experiments demonstrate that our framework on one hand well maintains the global facial structures, on the other hand recovers the detailed facial traits in high quality.

1. Introduction

For face identification, especially by human, it is desirable to render a high-resolution face image from the low-resolution one, which is called *face hallucination* or *face super-resolution*. A number of super-resolution techniques have been proposed in recent years [1][2][3][5][6][7][9][10]. They can be categorized into two representative approaches: the first is to produce a super-resolution image from a set of low-resolution ones based on learning [5][7][10]; another approach is to model the texture as a Markov Random Field (MRF) [3][6][7]. These methods are mainly applied to generic images. However, for face hallucination the utilization of the special properties of the faces is conductive to generate the high-resolution face images.

Baker et al. [1][2] develop a hallucination method under a Bayesian formulation. In the method, it infers the high frequency components from a parent structure with the assistance of training samples. Liu et al. [9] propose a two-step approach integrating a parametric global models with Gaussian assumption and linear inference and a nonparametric local model based on MRF. Both methods rely on explicit resolution-reduction-function, which is sometimes unavailable in practice. In [15], Wang and Tang develop an efficient face hallucination algorithm using an eigentransformation algorithm [12]. However, the method only utilizes global information without paying attention to local details. Inspired by a well-known manifold learning method, Locally Linear Embedding (LLE), Chang et al. [4] develop the Neighbor Embedding algorithm based on the assumption that the local distribution structure in sample space is preserved in the down-sampling process, where the structure is encoded by patch-reconstruction weights.

Face image is influenced by multiple factors. Modelling of the contribution of these factors is crucial to face analysis and synthesis. Recently, Vasilescu et al. introduce Multilinear Analysis to face modelling [13][14] and demonstrate its promising application in computer vision. In the method, equipped with tensor algebra, the multiple factors are unified in the same framework with the coordination between factors expressed in an elegant tensor product form.

In this paper, we propose a novel patch-based face hallucination framework, where the entire low-resolution image is partitioned into overlapped patches, and high-resolution patches are inferred respectively and then fused together to form a high-resolution image. For patch-based inference, two novel methods *TensorPatches* and *Coupled Residue Compensation* are integrated. First, TensorPatch model is developed to learn the relation of multiple constituent factors including identity and patch-locations, based on which, factor parameters can be solved to represent the local structure of the sample space. Based on the assumption that

the low-resolution space and high-resolution space share similar local distribution structure, the solved parameters are used for synthesizing high-resolution images. To further enhance the quality of the hallucinated image, we develop a coupled residue compensation algorithm based on a new statistical model called Coupled PCA. which infers the residual from low-resolution residue to high-resolution residue. Comparative experiments are conducted, which show the remarkable improvement achieved by our framework over other state-of-the-art algorithms.

2. TensorPatches

2.1. Theory of Multilinear Analysis

First of all, we briefly review the multilinear analysis framework[8][13][14]. The mathematical foundation of multilinear analysis is tensor algebra. *Tensor*, can be regarded as multidimensional generalization of conventional matrix. We denote an N^{th}-order tensor as $\mathcal{A} \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with each element denoted as $\mathcal{A}_{i_1 \cdots i_2 \cdots i_N}$ or $a_{i_1 \cdots i_2 \cdots i_N}$, where $1 \leq i_n \leq I_n$. In multilinear algebra terminology, each dimension of tensor is called a "*mode*". Mode-n vectors can be acquired by varying the nth-mode indices while keeping the indices of other modes fixed. All mode-n vectors constitutes a matrix, which is called *mode-n unfolding matrix* or *mode-n flattening matrix* of the tensor $\mathcal{A}$, denoted as $\mathbf{A}_{(n)} \in \mathbf{R}^{I_n \times (I_1 \cdots I_{n-1} I_{n+1} \cdots I_N)}$. The mode-$n$ vectors are just the column vectors of $\mathbf{A}_{(n)}$. The tensor can be restored by *mode-n folding* of the matrix $\mathbf{A}_{(n)}$.

Suppose we have a tensor $\mathcal{A} \in \mathbf{R}^{I_1 \times I_2 \times \cdots \times I_N}$, and a matrix $\mathbf{U} \in \mathbf{R}^{J_n \times I_n}$, the mode-$n$ product, denoted by $\mathcal{A} \times_n \mathbf{U}$, is also an N^{th}-order tensor $\mathcal{B} \in \mathbf{R}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$, whose entries are

$$(\mathcal{A} \times_n \mathbf{U})_{i_1 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_n} u_{j_n i_n}. \quad (1)$$

In addition, we have a generalized counter part of SVD for tensor, called high-order SVD (HOSVD), which was first proposed by Lathauwer et al. [8]

$$\mathcal{A} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n \cdots \times_N \mathbf{U}_N. \quad (2)$$

Here $\mathcal{C}$, is the core tensor satisfying sub-tensor-orthogonality and norm-ordering. $\mathbf{U}_1, \ldots, \mathbf{U}_N$ are all orthonormal matrices.

In a real application, the appearance variations of an visual object are always caused by diverse factors, like illumination, deformation and pose. Traditional linear methods, which treat factors of different aspects in additive manner, lack the ability of modelling the complexity of reality. As a generalization of linear analysis, multilinear analysis models the interaction between different factors in a modulating manner, and thus offers a more flexible and sophisticated framework for processing multi-factor problems.

In multilinear models, each factor has its own parameter space. The sample vector space and parameter spaces are connected by the following form

$$\mathcal{D} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n \cdots \times_N \mathbf{U}_N, \quad (3)$$

where $\mathcal{D}$ is an ensemble tensor obtained by grouping samples according to associated factors in all contributive aspects [13]. $\mathcal{C}$ is the core tensor which is responsible for controlling the interaction between different factor parameters. $\mathbf{U}_1, \cdots, \mathbf{U}_{N-1}$ are mode matrices capturing the variation patterns of different aspects. $\mathbf{U}_N$ is the prototype matrix storing the orthogonal vectors encoding the major variation patterns in the sample space which serve as bases for linear combination.

Eq (3) gives the representation for the whole ensemble in multilinear model. Followed this model, we deduce the representation for any individual sample as follows

$$\mathbf{x} = \mathcal{C} \times_1 \mathbf{v}_1^T \times_2 \mathbf{v}_2^T \times_3 \cdots \times_{N-1} \mathbf{v}_{N-1}^T \times_N \mathbf{U}_N. \quad (4)$$

2.2. Formulation of TensorPatches

Here we present a new mathematical approach, called *TensorPatches*, to explicitly account for local constituent factors of faces. TensorPatches applies multilinear image analysis and representation in a new way.

In our model, we employ a face database from m subjects with each facial image formed by a set of l small overlapped image patches. The number of pixels within each patch is assumed to be d.

Considering that each patch is affected by two major deterministic factors: the characteristics of the person and the position of the patch in the image, a 3^{rd} order tensor model is built to coordinate the interaction of these two constituent factors. The model is formulated as below

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{people} \times_2 \mathbf{U}_{patches} \times_3 \mathbf{U}_{pixels}, \quad (5)$$

where the data tensor $\mathcal{D}$ is the patch ensemble obtained by grouping the patches in training samples. $\mathcal{Z}$ governs the interaction between the two parametric factor space and the reconstruction bases. The two factor space are encoded in the two factor mode matrices: the $m \times m$ mode matrix $\mathbf{U}_{people}$ spanning the space of personal parameters and the $l \times l$ mode matrix $\mathbf{U}_{patches}$ spanning the space of position parameters, while the patch bases are stored in the $d \times d$ mode matrix $\mathbf{U}_{pixels}$ spanning the patch space.

In conclusion, the TensorPatch model explicitly accounts for the two major factors: identity and patch position. Besides, the variations yielded by inter-personal difference and position difference are combined in a mutual modulation manner, which is more powerful and flexible in modelling the reality.

2.3. Patch-based Reconstruction

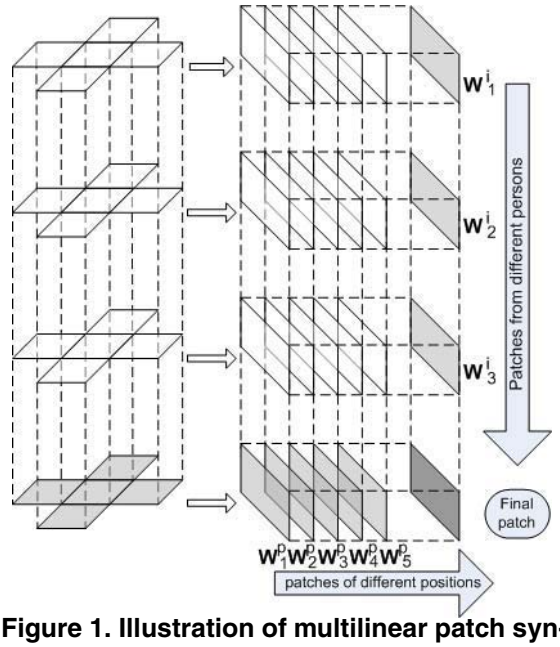


Figure 1. Illustration of multilinear patch synthesis. W^i_j are weights assigned to different persons, W^p_j are weights assigned to different positions.

Here we develop the algorithm for reconstruction of an individual patch. At first, we examine the samples in the training ensemble. For the jth patch of the ith individual, the $1 \times 1 \times d$ subtensor $\mathcal{P}_{i,j}$ is denoted as

$$\mathcal{P}_{i,j} = \mathcal{Z} \times_3 \mathbf{U}_{pixels} \times_1 \mathbf{v}_1^T \times_2 \mathbf{v}_2^T = \mathcal{TP} \times_1 \mathbf{v}_1^T \times_2 \mathbf{v}_2^T, \quad (6)$$

where $\mathbf{v}_1^T$ and $\mathbf{v}_2^T$ are the ith and jth row vector of mode matrix $\mathbf{U}_{people}$ and $\mathbf{U}_{patches}$, respectively. Thus an image patch $\mathbf{p}_{i,j}$ in the training ensemble corresponding to the ith person and jth position, can be extracted through mode-1(or 2) flattening the subtensor $\mathcal{P}_{i,j}$ as

$$\mathbf{p}_{i,j} = (\mathbf{P}_{i,j})^T_{(1)} = (\mathbf{P}_{i,j})^T_{(2)}. \quad (7)$$

Generalization of the analysis above follows that for any patch with identity-related parameter $\mathbf{v}_1$ and position-related parameter $\mathbf{v}_2$, it can be obtained by tensor product in Eq.(6) ensued by flattening.

Considering that the row vectors capture the major variation patterns in two parametric spaces, we can assume that the parameter vector $\mathbf{v}_n$ exists in the row space of the mode matrix $\mathbf{U}_n$. Thus, $\mathbf{v}_n$ can be represented as a linear combination of row vectors of $\mathbf{U}_n$, i.e. $\mathbf{v}_n^T = \mathbf{w}_n^T \mathbf{U}_n$.

Motivated by [4], each patch should be reconstructed using the nearest neighbors. In our framework, the neighborhood is defined as the patches residing in the adjacent positions and from the training images containing the most similar patches in these positions. Concretely, we set N_1 as indices of nearest persons and N_2 as indices of adjacent positions for each patch. Thus, for each patch, patches from m nearest persons in the training set and l adjacent positions are selected for reconstruction.

Then we can rewrite Eq.(6) employing tensor property

$$\begin{aligned} \mathcal{P} &= \mathcal{TP} \times_1 \mathbf{w}_1^T \tilde{\mathbf{U}}_1 \times_2 \mathbf{w}_2^T \tilde{\mathbf{U}}_2 \\ &= (\mathcal{TP} \times_1 \tilde{\mathbf{U}}_1 \times_2 \tilde{\mathbf{U}}_2) \times_1 \mathbf{w}_1^T \times_2 \mathbf{w}_2^T \\ &= \widetilde{\mathcal{TP}} \times_1 \mathbf{w}_1^T \times_2 \mathbf{w}_2^T, \end{aligned} \quad (8)$$

where $\tilde{\mathbf{U}}_n$ is a partial matrix of mode matrix $\mathbf{U}_n$, denoted as $\tilde{\mathbf{U}}_n = \mathbf{U}_n(N_n)$ and N_n is the indices set containing row indices corresponding to nearest patches. A new Tensor-Patches $\widetilde{\mathcal{TP}}$ is sub-TensorPatches, which is function of two index set N_1 and N_2, denoted as $\mathcal{TP}(N_1, N_2)$.

Moreover, for consistency and stability, we impose the normalized constraints $\sum_{i=1}^{m} w_{1i} = 1$ and nonnegative constraint $\sum_{j=1}^{l} w_{2j} = 1, \quad w_{2j} \geq 0$ on the weight vectors.

In a real synthesis application, for a new patch sample $\mathbf{p}$ (d-dimensional column vector) with its parameters unknown, we are required to first extract the parameters before reconstruction. Before developing the scheme for solving parameters for a new patch sample, we examine the simplest case where only mode-1 factor parameter (people identity) is unknown, i.e. the vector $\mathbf{w}_1$ is unknown. Then we can rewrite the formula (6) by applying commutative law

$$\mathcal{P} = (\widetilde{\mathcal{TP}} \times_2 \mathbf{w}_2^T) \times_1 \mathbf{w}_1^T = \mathcal{S}_1 \times_1 \mathbf{w}_1^T, \quad (9)$$

where $\mathcal{S}_1$ is an N^{th}-order tensor called mode-1 base tensor. It can be seen that all samples with the same factor (except for mode-1 factor) constitutes the space spanned by mode-n vectors in $\mathcal{S}_1$. Look at it from another perspective, we can convert the expression in matrix form as follows

$$\mathbf{P}_{(1)} = \mathbf{w}_1^T (\mathbf{S}_1)_{(1)} \Longrightarrow \mathbf{p} = \mathbf{P}^T_{(1)} = (\mathbf{S}_1)^T_{(1)} \mathbf{w}_1. \quad (10)$$

Similarly, when parameters of $n(n = 1, 2)$th mode are unknown, we have

$$\mathbf{p} = (\mathbf{S}_n)^T_{(n)} \mathbf{w}_n, \quad (11)$$

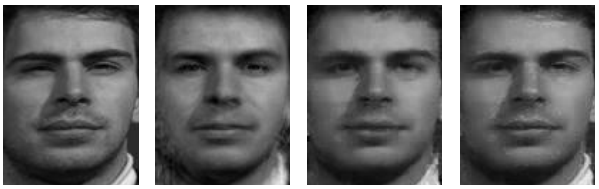
here, the multilinear model degenerates to linear model, then we can solve the parameter vector in that mode by constraint least square with the solution denoted as $\mathbf{w}_n = lsq((\mathbf{S}_n)_{(n)}, \mathbf{p})$.

Then for the general case where **all parameters are unknown**, we can solve all of them by a scheme called *Constraint Alternate Least Square* as follows ($k = 1, 2, \cdots$)

$$\begin{aligned} \mathcal{S}_1^{(k+1)} &= \widetilde{\mathcal{TP}} \times_2 (\mathbf{w}_2^{(k)})^T \\ \mathbf{w}_1^{(k+1)} &= lsq((\mathbf{S}_1)^{(k+1)}_{(1)}, \mathbf{p}) \\ \mathcal{S}_2^{(k+1)} &= \widetilde{\mathcal{TP}} \times_1 (\mathbf{w}_1^{(k)})^T \\ \mathbf{w}_2^{(k+1)} &= lsq((\mathbf{S}_2)^{(k+1)}_{(2)}, \mathbf{p}). \end{aligned} \quad (12)$$

Denote the optimal parameter vectors obtained in the final step as $\mathbf{w}_1^*, \mathbf{w}_2^*$. Then the reconstructed version of an image patch $\mathbf{p}$ can be acquired by $\widetilde{\mathcal{TP}} \times_1 \mathbf{w}_1^{*T} \times_2 \mathbf{w}_2^{*T}$.

Figure 2 illustrates the reconstruction results compared to traditional linear reconstruction methods. Multilinear patch-based reconstruction on one hand maintains the local features well, on the other hand basically keeps the inter-patch continuity, thus improves the smoothness of the image.



(a) Original (b) PCA (c) Local linear (d) Multilinear

Figure 2. Comparative results of reconstruction

2.4. TensorPatch Super-Resolution

As in Locally Linear Embedding(LLE), our hallucination algorithm is based on the assumption that small patches in low resolution space and high resolution space form manifolds with the same local structure, which is characterized by the weights of neighboring patches. Accordingly we can synthesize the high resolution patch employing the weights infered from input low resolution patches.

In the training stage, two TensorPatch models are trained on the low- and high-resolution image patch sets, denoted as $(\{\mathbf{p}_{L,j}^{(i)} |\ \mathbf{p}_{L,j}^{(i)} \in I_L^{(i)}, 1 \le i \le m, 1 \le j \le l\}$ and $\{\mathbf{p}_{H,j}^{(i)} |\ \mathbf{p}_{H,j}^{(i)} \in I_H^{(i)}, 1 \le i \le m, 1 \le j \le l\})$ respectively.

In the testing stage, we divide the whole image into a set of small overlapped image patches as elements for hallucination. For each patch, we first analyze each input low-resolution patch using the low-resolution-TensorPatches to produce the local weight vectors. Secondly the corresponding high-resolution patch is rendered by high-resolution-TensorPatches based synthesis using the learned parameters. To assure local compatibility and smoothness between the hallucinated patches, we superpose patches in adjacent regions of one image and blend pixels in the overlapping area.

The TensorPatch super-resolution algorithm is summarized as follow:

- **Step1**. For each patch $\mathbf{p}_{L,j}$ $(j = 1, 2, \cdots, l)$ in an input low-resolution face I_L:
  - $(\mathbf{a})$ Find K_1-NNs of different people in the training patch ensemble $\{\mathbf{p}_{L,j}^{(i)}\}_{i=1}^m$, and find K_2-NNs for adjacent patches in the input image $\{\mathbf{p}_{L,t}\}_{t=1}^l$. Denote them as N_{people} and $N_{patches}$.
  - $(\mathbf{b})$ Compute the multilinear reconstruction weights $\mathbf{w}_1$ and $\mathbf{w}_2$ based on the sub-TensorPatches $\mathcal{TP}_L(N_{people}, N_{patches})$ using ALS procedure introduced in the previous subsection.
  - $(\mathbf{c})$ Synthesize the high-resolution patch $\mathbf{p}_{H,j}$ from mode-1 unfolding the high-resolution tensor $\mathcal{TP}_H(N_{people}, N_{patches}) \times_1 \mathbf{w}_1 \times_2 \mathbf{w}_2$.
- **Step2**. Concatenate and integrate the hallucinated high-resolution patches $\{\mathbf{p}_{H,j}\}_{j=1}^l$ to form a facial image, which is the target high-resolution facial image, with local compatibility and smoothness constraints.

The proposed algorithm can recover the global face structure and main local features of the target high-resolution face in the super-resolution image. We denote the hallucinated result of TensorPatches as I_H^{TP}.

3. Coupled Residue Compensation

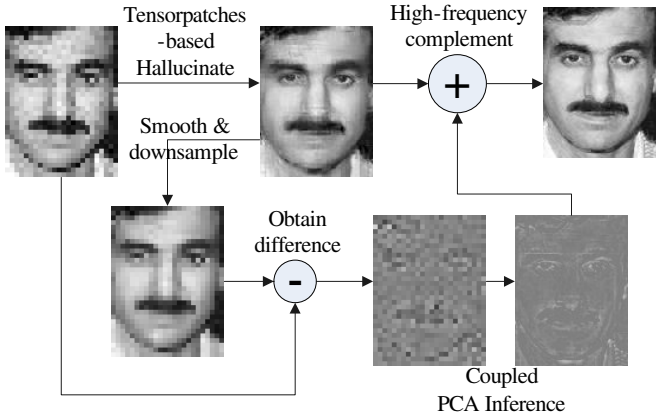


Figure 3. Architecture of Coupled Residue Compensation.

In the TensorPatch model, by considering both identity and adjacent patches, multilinear synthesis can achieve fairly good approximation. However, there still exist some difference between down-sampled version of hallucinated image and the original low-resolution image. Inspired by the observation, we develop a scheme *Coupled Residue Compensation*, which utilizes the difference to further enhance the hallucination quality by residue compensation. Our residue compensation algorithm is based on a novel statistical model named *Coupled PCA* developed by us, which is employed to learn the relation between the low-resolution residue space and the high-resolution residue space. The

whole architecture of Coupled Residue Compensation is shown in Figure 3.

3.1. Coupled PCA

In order to establish the relationship between two spaces, we develop a model named Coupled PCA, which connects two spaces by introducing a hidden space.

Suppose we have two spaces Ω_X and Ω_Y. We use $\mathbf{x}$ to denote the vector in space Ω_X, and use $\mathbf{y}$ to denote the vector in space Ω_Y. Assume there exists a linear relation between the two spaces, then we can approximate $\mathbf{y}$ by $\mathbf{Ax}$.

In statistics, multivariate regression is available for linking two linear spaces. The multivariate linear regression problem can be formulated as the following least-square optimization problem:

$$\mathbf{A} = \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{A}\mathbf{x}_i\|^2. \tag{13}$$

It can be easily shown that the solution is

$$\mathbf{A} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}, \tag{14}$$

here, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$.

However, in real application, it is always the case that samples in Ω_X or Ω_Y only distribute in subspaces capturing the intrinsic variations shared by both low- and high-resolution patch spaces. To effectively and robustly evaluate the linear relation, we integrate the multivariate linear regression and the well-known dimension-reduction method PCA by explicitly introducing the concept *Common Hidden Space*.

In our formulation, the vector $\mathbf{x}_i$ in space Ω_X and the vector $\mathbf{y}_i$ in space Ω_Y correspond to the same vector $\mathbf{h}_i$ in the hidden space. In ideal case without noise, we have

$$\mathbf{x}_i = \mathbf{B_X}\mathbf{h_i} \qquad \mathbf{B_X} \in \mathbf{R}^{d_x \times d_h}, \tag{15}$$

$$\mathbf{y}_i = \mathbf{B_Y}\mathbf{h_i} \qquad \mathbf{B_Y} \in \mathbf{R}^{d_y \times d_h}. \tag{16}$$

The dimension of common hidden space is smaller than that of space Ω_X and space Ω_Y: $d_h < d_x, d_h < d_y$. In addition, the base matrices $\mathbf{B_X}$ and $\mathbf{B_Y}$ are orthogonal matrices. Then the linear relation between two spaces is as below

$$\mathbf{y} = \mathbf{B_Y}\mathbf{B_X}^T\mathbf{x}. \tag{17}$$

Combining the formula (17) and formula (13), the Coupled PCA can be finally formulated as minimization of the following error energy function

$$\mathrm{E}(\mathbf{B_X}, \mathbf{B_Y}) = \sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{B_Y}\mathbf{B_X}^T\mathbf{x}_i\|^2. \tag{18}$$

The problem can be solved by Alternate Least Square (ALS), where we alternately optimize $\mathbf{B_X}$ and $\mathbf{B_Y}$ with the other matrix fixed.

In summary, traditional multivariate regression model is suitable for analyzing the coupled linear relation between two spaces, however, it is unstable and inefficient for high-dimensional spaces such as images; while standard PCA offers a powerful approach to handle high-dimensional data but lacks the capability of processing the coupled problem. Our scheme integrates the advantages of both multivariate regression and PCA to give a potent solution for analyzing high-dimensional coupling spaces.

3.2. Residue Compensation using Coupled PCA

It is observed that there are still differences between reconstructed patches and groundtruth patches. It is desirable to utilize the information in low-resolution residue to further enhance the quality of high-resolution hallucinated images by residue compensation.

In our face hallucination algorithm, the Coupled PCA is applied to establish the relation between low-resolution patch-residue and high-resolution patch-residue. In the training stage, we hallucinate high-resolution patches from a set of low resolution patches. Then low- and high-resolution residue for each patch sample can be constructed as below: one is obtained by subtracting the input low resolution image by a down-sampled version of the hallucinated image; the other one is obtained by subtracting the groundtruth high resolution image by hallucinated image. Then Coupled PCA can be trained on these two spaces spanned by low- and high-resolution patch residue. In testing stage, the high resolution patch-residue can be inferred by low-resolution patch-residue and compensated to the hallucinated image.

The process can be briefly described as below:

1. For an input low resolution image, the super resolution image I_H^{TP} is hallucinated by TensorPatch algorithm.

2. Construct the low resolution difference image I_L^R by subtracting the input image I_L with down-sampled version of the hallucinated image I_H^{TP}.

3. Infer the super resolution residue I_H^R by formula (17).

4. Add the inferred residue image I_H^R to the super resolution image I_H^{TP} hallucinated by TensorPatches to acquire the final result $I_H^* = I_H^{TP} + I_H^R$.

Experiments show that using Coupled Residue Compensation, the hallucination quality can be notably improved.

4. Experiments

Our experiments were conducted with a subset of *FERET* dataset [11], which consists of about 1400 images. Among all these samples, we select 600 samples for training

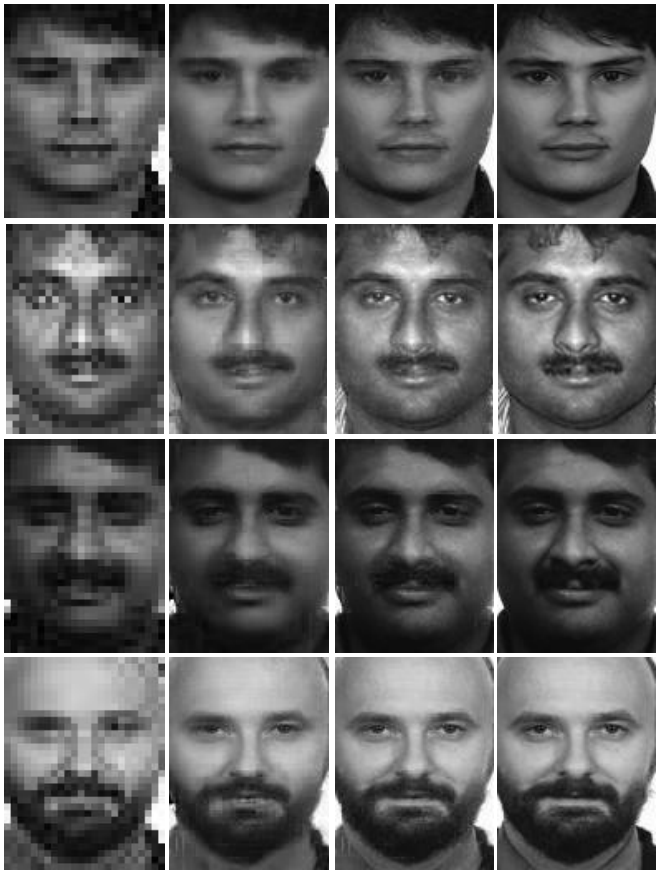


Figure 4. The hallucinated results. (a) is the low-resolution 24×32 input. (b) is the TensorPatch hallucinating result I_H^{TP}. (c) is the final result $I_H^* = I_H^{TP} + I_H^R$. (d) is the original high-resolution 96×128 image.

the TensorPatch model and another 600 samples for training the Coupled PCA for residue compensation. Other samples are for testing. As a necessary preamble steps, we perform geometric normalization by affine transform based on coordinates of eyes and mouth. After the transform, each image is cropped to a canonical 96×128 grayscale image as the high resolution one. The corresponding low-resolution images are obtained by smoothing and down-sampling, which are 24×32 images.

In our experiments, we first train the TensorPatch model. For each low resolution image, 682 overlapped patches are extracted by sliding a small 3×3 window pixel by pixel. For high resolution images, 682 12-by-12 patch are extracted as well. The patches in low resolution image and high resolution image are in one-to-one correspondence. The training process of TensorPatch model eventually gives two TensorPatches: $\mathcal{TP}_L \in \mathbf{R}^{600 \times 682 \times 9}$ for low resolution images and $\mathcal{TP}_H \in \mathbf{R}^{600 \times 682 \times 144}$ for high resolution images.

In the testing stage of TensorPatch hallucination, each patch is hallucinated in analysis-by-synthesis manner. By a series of experiments, we choose $m = 5$ and $l = 9$ for K-NN schemes. That is, only the patches from the 5 nearest persons and in the same position or 8 neighboring overlapped positions are taken into account. And the number of iterations for each ALS procedure is set to 5. Our experiments show that such configuration on parameters yields the most satisfactory results.

Further, a second set containing 600 samples, which is disjoint from the training set for TensorPatches, is used for training the Coupled PCA for residue compensation.

The resultant images are shown in Figure 4. We can see that only the TensorPatches can produce good hallucinated results, and the coupled compensation stage further enhances the quality, which makes the final image approximate the groundtruth fairly well.

We compare our algorithm with other representative ones, including Cubic B-Spline, Baker's and Ce Liu's algorithms, in Figure 5. The quality of cubic B-Spline reconstruction is rather rough. Baker et al's method produces a blurred face with important features lost. Liu et al's method though can hallucinate a super-resolution image of acceptable quality, some subtle characteristics are blurred. Our method has advantages over others in terms of preserving both global structure and subtle details.

5. Conclusions

We have shown that the face hallucination framework integrating TensorPatch model and Coupled Residue Compensation is capable of producing high quality super-resolution images. Compared to other approaches, our method has the advantages in preserving global structure and local detail features by seamless incorporating multiple factors and constraints. Our experiments clearly show the efficacy of the new models.

Acknowledgement

The authors thank Zhifeng Li for his constructive suggestion of applying Tensor algebra to the super-resolution problems. The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region.

References

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

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

(a) Input 24*32 (b) Cubic B-Spline (c) Baker et al. (d) Liu et al. (e) Our method (f) Original 96*128

Figure 5. Comparison between our method and others.

[3] J.D. Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," in *Proc. of SIGGRAPH*, pp. 361-368, 1997.

[4] H. Chang, D.Y. Yeung, and Y. Xiong, "Super-Resolution Through Neighbor Embedding," in *Proc. of CVPR*, Vol. 1, pp. 275-282, 2004.

[5] M. Elad and A. Feuer, "Super-Resolution Reconstruction of Image Sequences," *IEEE Trans. on PAMI*, 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] L.D. Lathauwer, B.D. Moor, and J. Vandewalle, "Multilinear Singular Value Tensor Decompositions," *SIAM Journal on Matrix Analysis and Applications*, Vol. 21, No. 4, pp. 1253-1278, 2000.

[9] 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.

[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] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. on Circuits, Systems, and Video Technology*, Vol. 14, No. 1, pp. 50-57, 2004.

[13] M.A.O. Vasilescu and D. Terzopoulos, "Muiltilinear Analysis of Image Ensembles: TensorFaces," in *Proc. of ECCV*, pp. 447-460, 2002.

[14] M.A.O. Vasilescu and D. Terzopoulos, "Multilinear Subspace Analysis of Image Ensembles," in *Proc. of CVPR*, Vol. 2, pp. 93-99, 2003.

[15] 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.