# Learning-based Face Hallucination in DCT Domain

Wei Zhang    Wai-Kuen Cham
Department of Electronic Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
{zhangwei,wkcham}@ee.cuhk.edu.hk

## Abstract

*In this paper, we propose a novel learning-based face hallucination framework built in DCT domain, which can recover the high-resolution face image from a single low-resolution one. Unlike most previous learning-based work, our approach addresses the face hallucination problem from a different angle. In details, the problem is formulated as inferring DCT coefficients in frequency domain instead of estimating pixel intensities in spatial domain. Experimental results show that DC coefficients can be estimated fairly accurately by simple interpolation-based methods. AC coefficients, which contain the information of local features of face image, cannot be estimated well using interpolation. We propose a method to infer AC coefficients by introducing an efficient learning-based inference model. Moreover, the proposed framework can lead to significant savings in memory and computation cost since the redundancy of the training set is reduced a lot by clustering. Experimental results demonstrate that our approach is very effective to produce hallucinated face images with high quality.*

## 1. Introduction

As an active research field in computer vision, super-resolution is to produce high-resolution image (HRI) or frames from low-resolution image (LRI) or frames. Recently, an interesting topic within super-resolution, *face hallucination*, has aroused much attention. This term is firstly introduced by Baker and Kanade [1], whose particular interest is to generate a high-resolution face image from low-resolution input. It can be widely applied in many fields ranging from image compression to face identification. Especially in video surveillance, a higher resolution face image with detailed facial features will be obviously significant to raise the system's performance.

### 1.1. Previous Work

Face hallucination from a single low-resolution face image which is also referred as single-image super-resolution problem received a lot of attention in recent years. A number of related super-resolution and face hallucination algorithms have been proposed, which can be grouped into three types. Interpolation-based algorithms (e.g. Bilinear, Cubic B-Spline) suffer from severe blurring problem especially when the resolution of the input is very low. Reconstruction-based methods [3] [4], which try to model the process of image formulation to build the relationship between LRI and HRI based on reconstruction constraints and smoothness constraints, are quite limited by the number of input LRIs and usually cannot work well in single-image super-resolution problem.

Recently, learning-based methods become very popular. Usually, the unknown HRI is inferred by making use of some training set directly or indirectly. Compared with other methods, learning-based method can achieve higher magnification factor and output better results especially for single-image super-resolution problem [5]. Baker and Kanade [1] [2] presented a pioneering work on hallucinating face image based on a Bayesian formulation. The target HRI is inferred by resorting to a training set. Freeman et al. [6] proposed a well-known parametric Markov network to learn the statistics between unknown *scene* and observed *image*. This framework was applied to super-resolution problem as well as some other low-level vision problems. Such Markov network was extended and adopted by Sun et al. [7], Bishop et al. [8], Liu et al. [9] and Wang et al. [10]. For instance, Liu et al. [9] developed a two-step statistical modeling approach for face hallucination which integrates a global parametric model and a local nonparametric model. Wang et al. [10] proposed a combination model by integrating the super-resolution constraint and the patch based image co-occurrence constraint for super-resolution problem. Besides, Liu et al. [11] hallucinated the low-resolution face image by introducing a TensorPatch model and then devised a residue compensation step to enhance the hallucination result. All above mentioned learning-based methods

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

are built in spatial domain for the inference of pixel intensities of the target HRI, and differed with each other on the learning manner from the training set. A major problem of these methods is the high computation requirement due to the complex learning process. Especially in the Markov network based inference model, rather taxing computation and heavy memory load are required when the training set becomes very large. Very recently, some transform domain based methods are presented. Tuan et al. [12] implemented the prevalent Markov-based work [6] in DCT domain for fast super-resolving the compressed video. Karl et al. [13] applied support vector regression (SVR) to super-resolution and utilized DCT structural properties to aid in solving their proposed regression structure.

### 1.2. Our Method

In this paper, we propose an efficient learning-based face hallucination framework built in the Discrete Cosine Transform (DCT) domain which is shown in Figure 1. More specifically, instead of estimating pixel intensities directly as the traditional learning-based algorithms, we concern ourselves with inferring the DCT coefficients, which contains two parts: DC coefficients estimation and AC coefficients inference. DC coefficient, which represents the average energy of a target block, can be estimated fairly accurately by some simple interpolation-based methods (e.g. Bilinear, Cubic B-Spline). AC coefficients, which contain the information of local features such as edges and corners around eyes, mouth of face image, cannot be estimated well by interpolation. Therefore, a simplified learning-based inference model is proposed to tackle this challenging problem. The basic idea of our method is that we are only interested in learning the local facial features embodied in AC coefficients from a specific training set, since common facial features are very similar and can be shared more easily in different types of faces. Thus, on the one hand, a more specific and efficient training set for AC coefficient priors can be built and used, on the other hand, without considering DC coefficients, our learning process will be more robust since it is much less influenced by image illumination. Moreover, in order to make our method more efficient and reduce the redundancy of the training set, a compact block dictionary is built by a clustering-based training scheme as introduced in Section 5.

Particularly, the intermediate hallucinated results $I_H$ in Figure 1 is defined as a preprocessed image by a prefiltering scheme [14] [15] which processes the block boundaries to remove the correlation of neighboring blocks. Then a reasonable assumption can be made that each HRI block in our AC coefficient inference model is independent with its adjacent HRI blocks according to the analysis of AC coefficient correlation. This is why our inference model is much simplified compared with the common used Markov network. Finally, the output $I_H^*$ can be obtained from $I_H$ by postfiltering, which is the inverse of the prefiltering. Besides, inspired by Locally Linear Embedding (LLE) [17] [18], a more general way of utilizing training priors is adopted in our learning process. In details, each target HRI block in our inference model is derived from multiple nearest training samples instead of only one.

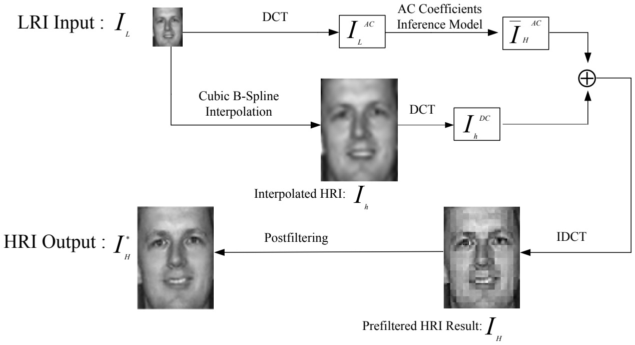


Figure 1. The proposed face hallucination framework.

The rest of this paper is organized as follows. Section 2 is the problem formulation and an overview of the proposed work. The simplified AC coefficients inference model is introduced in Section 3. The reconstruction of the target HRI is given in Section 4. The clustering-based training scheme is stated in Section 5. Experimental results are discussed in Section 6. Section 7 gives some concluding remarks.

## 2. Problem Formulation and Overview of the Proposed Framework

Face hallucination from a single low-resolution face image is a typical ill-posed problem. Usually, the inferred HRI will suffer from blurring problem especially in the local details (e.g. edges, corners). For example in Figure 2, Cubic B-Spline interpolation is used to enlarge a $24 \times 32$ low-resolution face image to $96 \times 128$ high-resolution image. The difference image shown in Figure 2 (d) shows that Cubic B-Spline works well in the smooth parts of face. But lot of details are lost in the parts (e.g. eyes, mouth and nose) with abundant local features. In frequency domain, this is because of the great loss of AC coefficients which contain the information of local details. Therefore, the challenging problem in face hallucination is to infer enough AC components which can make the reconstructed HRI remain sharp in local features.

In the proposed framework shown in Figure 1, face hallucination is treated as inferring DC and AC coefficients for each block of the prefiltered HRI $I_H$. Such formulation will benefit us in several aspects:

1. DC coefficient which represents the average energy of a target block, can be estimated fairly accurately by

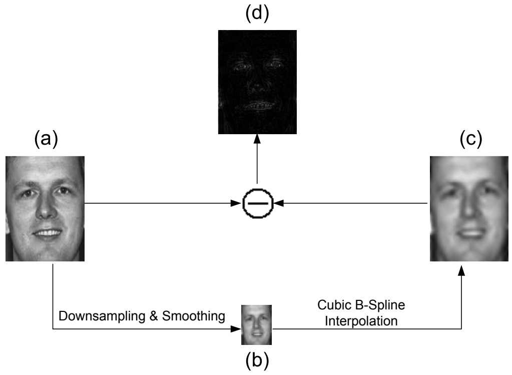


Figure 2. Face Hallucination using Cubic B-Spline Interpolation. (a) original HRI (96 × 128); (b) input LRI (24 × 32); (c) interpolated HRI using Cubic B-Spline; (d) difference image.

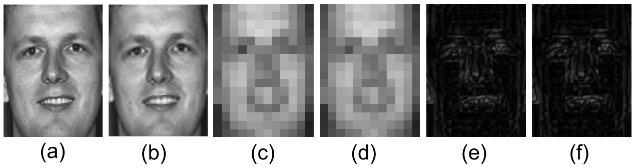


Figure 3. Image coded by 8 × 8 DCT. (a) original image (96 × 128); (b) reconstructed image with 16 coefficients (1 DC plus the first 15 AC coefficients in zig-zag order); (c) reconstructed image only with the original DC coefficients; (d) reconstructed image only with the DC coefficients which are estimated by Cubic B-Spline interpolation; (e) reconstructed with all 63 AC coefficients (absolute image); (f) reconstructed image with only the first 15 AC coefficients (absolute image).

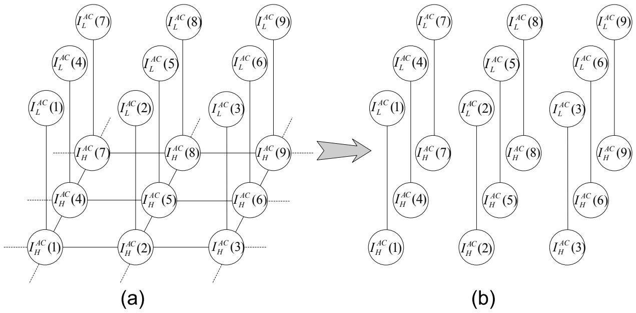


Figure 4. Graphical model for AC coefficients inference. (a) Markov network[6]; (b) our simplified inference model.

some interpolation-based methods such as Cubic B-Spline. This can be demonstrated by comparing Figure 3 (c) and (d).

2. As shown in Figure 3 (e), we only need to focus on building a specific learning-based inference model for AC coefficients which correspond to the local details of face image. The basic idea is that we are only interested in learning the local facial features (e.g. edges, corners around eyes, mouth, nose) from a specific training set.

3. A simplified learning-based inference model can be developed to infer AC coefficients efficiently based on a reasonable assumption that blocks of the prefiltered HRI built in DCT domain are independent with each other.

4. The data dimension of training and testing set can be reduced a lot. For example, the image shown in Figure 3 (a) is encoded by $8 \times 8$ DCT. Figure 3 (b) shows the reconstructed image from 16 DCT coefficients (1 DC plus the first 15 AC coefficients in zig-zag order) by Inverse DCT (IDCT). It can be observed that Figure 3 (b) are very similar with the original image and contains most of local features of a face image, although only a small part of coefficients is used in the decoding. This inspired us it is not necessary to infer all AC coefficients as shown in Figure 3 (e) for each block of the target $I_H$. For $8 \times 8$ DCT, the first 15 AC coefficients are enough to produce a satisfying result with detailed local features as shown in Figure 3 (f). So the dimension of HRI block can be reduced from 64 in spatial domain to 15 in DCT domain in this case. This will make our learning process much faster than traditional learning-based methods built in spatial domain. Moreover, it will save a lot of computer memory which can allow us to collect more training samples and thus be less bounded by the limits of learning-based algorithm [5].

In conclusion, as shown in Figure 1, the proposed framework can be divided into two steps: firstly, infer the prefiltered HRI $I_H$ in DCT domain, which contains two parts: AC coefficients inference and DC coefficients estimation; Secondly, reconstruct the final hallucinated results $I_H^*$ from the prefiltered $I_H$ by postfiltering.

## 3. Learning-based AC Coefficients Inference Model

The goal of this part is to infer AC coefficients $I_H^{AC}$ for the target HRI given the input LRI AC coefficients $I_L^{AC}$. It can be formulated as a problem of finding the optimal $I_H^{AC}$ that can maximize the posterior probability:

$$I_H^{AC\,*} = \arg\max_{I_H^{AC}} \; p(I_H^{AC}|I_L^{AC}) \qquad (1)$$

As shown in Figure 4 (a), a typical Markov network [6] of low-level vision field can be adopted to formulate the above optimization problem. Node $I_H^{AC}(i)$ and node $I_L^{AC}(i)$ are used to represent unknown $ith$ high-resolution block of HRI and the observed $ith$ low-resolution block of LRI respectively. These links between nodes indicate statistical dependencies. So the MRF model in Figure 4 (a) implies two things: 1) HRI block $I_H^{AC}(i)$ provides all the information about the observed LRI block $I_L^{AC}(i)$, since $I_H^{AC}(i)$ has the only link to $I_L^{AC}(i)$; 2) HRI block $I_H^{AC}(i)$ gives information about adjacent HRI blocks by the links from $I_H^{AC}(i)$ to adjacent HRI blocks.

Since $p(I_H^{AC}|I_L^{AC}) = \frac{p(I_H^{AC}, I_L^{AC})}{p(I_L^{AC})}$ and $p(I_L^{AC})$ is constant over $I_H^{AC}$, Eq.(1) can be rewritten as

$$I_H^{AC\,*} = \arg\max_{I_H^{AC}} p(I_H^{AC}, I_L^{AC}) \qquad (2)$$

According to the MRF model in Figure 4 (a), the joint probability of $I_L^{AC}$ and $I_H^{AC}$ can be decomposed as:

$$p(I_H^{AC}, I_L^{AC}) = p(I_H^{AC}(1), ..., I_H^{AC}(n), I_L^{AC}(1), ..., I_L^{AC}(n))$$
$$= \frac{1}{Z} \prod_{(i,j)} \psi(I_H^{AC}(i), I_H^{AC}(j)) \prod_i \phi(I_H^{AC}(i), I_L^{AC}(i)) \qquad (3)$$

Where $Z$ is a normalization constant factor, $n$ denotes the number of block pairs, $(i, j)$ indicates neighboring blocks. Both $\psi$ and $\phi$ are introduced pairwise compatibility functions which model the two kinds of dependencies in Figure 4 (a). They are learned from the training set.

Now the optimization problem of Eq.(1) becomes:

$$I_H^{AC\,*} =$$
$$\arg\max_{I_H^{AC}} \prod_{(i,j)} \psi(I_H^{AC}(i), I_H^{AC}(j)) \prod_i \phi(I_H^{AC}(i), I_L^{AC}(i)) \qquad (4)$$

By using the loopy Belief Propagation (BP) algorithm [6], the target $I_H^{AC}$ can be inferred from a training set based on Eq.(4). However, the product of these terms in Eq.(4) is expensive to evaluate, meaning that attaining a global optimum will be difficult and certainly time consuming. In order to render the inference model as well as the optimization more tractable, let's firstly make a brief analysis on the correlation among AC coefficients.

## 3.1. Analysis of AC Coefficient Correlation

As a popular block transform, DCT refers to a separable orthogonal or nearly orthogonal linear mapping of blocks of image pixels into blocks of transform coefficients. For example, given a $KM \times LM$ image, $M \times M$ DCT will map it into a $K \times L$ grid of $M \times M$ coefficient blocks. In this way, the image can be represented with DCT coefficients by two forms: block representation and subband representation [15]. Figure 5 shows an example when $M = 4, K = L = 2$. Each block $(x, y)$ $(0 \le x < K, 0 \le y < L)$ gathers all coefficients at the same spatial location $(x, y)$ from every subband, represents different frequency components of a local spatial region. The coefficient $(u, v)$ $(0 \le u, v < M)$ is located at position $(x, y)$ in subband $(u, v)$. So, subband $(u, v)$ collects all coefficients at $(u, v)$ from every block. Figure 5 shows that every coefficient in each block has two kind of neighbors: block neighbors and subband neighbors. As a result, there are two kinds of correlation for each AC coefficient.

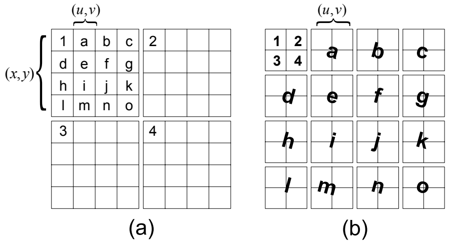


Figure 5. Block representation (a) and subband representation (b) for $4 \times 4$ DCT coefficients.

From the view of block representation as shown in Figure 5 (a), each AC coefficient is highly uncorrelated with its block neighbors because it is surrounded by these AC coefficients corresponding to different orthogonal or nearly orthogonal subbands. Therefore, this correlation is very weak and can be ignored.

From the view of subband representation as shown in Figure 5 (b), each AC coefficient is surrounded by its subband neighbors. Since each subband contains a part of the global information of the image, this correlation referred as interblock correlation is stronger than the last correlation. But compared with the correlation of neighboring blocks in spatial domain, it can also be regarded as neglectable. In fact, this correlation is not considered too in most block coding algorithms such as JPEG [16].

Besides, in order to avoid the blocking artifacts which often occur in block-based technology, a prefiltering scheme [14] [15] as shown in Figure 6 (a) is performed block-wise locally along the the block boundaries to remove the correlation between neighboring blocks in spatial domain. For instance, prefilter $P$, depicted in Figure 7 (a), is performed in a separable fashion similar with DCT to remove the $8 \times 8$ block neighboring correlation of the training HRI priors by the prefiltering scheme. As stated in Section 2, the intermediate result $I_H$ of the proposed framework is defined as the prefiltered HRI. Given $I_H$, postfilter $P^{-1}$ depicted in Figure 7 (b) as the inverse of the prefilter, is adopted to reconstruct the final HRI result $I_H^*$ by the postfiltering scheme as shown in Figure 6 (b).

Now we get a conclusion that each AC coefficient of the prefiltered image can be assumed to be neither corre-

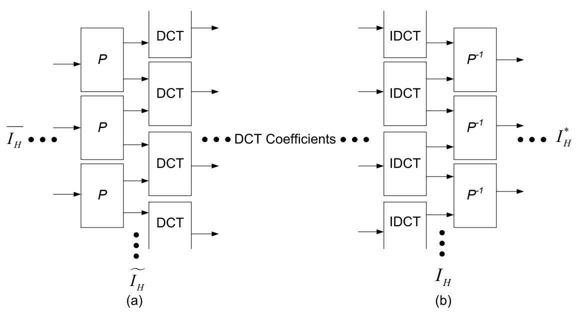


Figure 6. Prefiltering (a) and Postfiltering (b) are performed along the block boundaries block-wise locally similar with DCT. (a). Prefilter $P$ is adopted to preprocess the training HRI priors $\overline{I_H}$ as prefiltered HRI priors $\widetilde{I_H}$. (b). Postfilter $P^{-1}$ is adopted to reconstruct the final hallucinated result $I_H^*$ from the intermediate result $I_H$ which is defined as prefiltered HRI at the beginning.

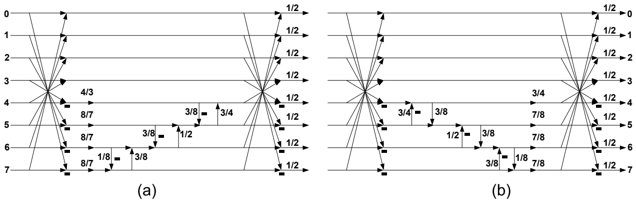


Figure 7. Prefilter $P$ (a) and Postfilter $P^{-1}$ (b) for 8 × 8 block processing.

lated with its block neighbors nor correlated with its subband neighbors. As a result, a reasonable assumption can be made that each block $I_H^{AC}(i)$ in the target prefiltered HRI is independent with its adjacent HRI blocks.

## 3.2. A Simplified AC Coefficients Inference Model

Since each block $I_H^{AC}(i)$ of the target prefiltered $I_H$ is independent with its neighboring HRI blocks, the Markov network as shown in Figure 4 (a) can be simplified a lot by taking out of all links among HRI blocks. Namely, in our AC coefficients inference model as shown in Figure 4 (b), it is not necessary to consider the compatibility function $\psi$ which models the dependencies among HRI blocks. So Eq.(4) can be simplified as:

$$I_H^{AC\,*} = \underset{I_H^{AC}}{\arg\max} \prod_i \phi(I_H^{AC}(i), I_L^{AC}(i)) \tag{5}$$

Thus the next problem is to build a reasonable compatibility function $\phi(I_H^{AC}(i), I_L^{AC}(i))$ for our inference model.

### 3.2.1 Compatibility Function Building based on Locally Linear Embedding (LLE)

LLE [17] is a promising manifold learning method which is to map high dimensional data into a low dimensional space by preserving the neighborhood relationship. This theory was extended in some super-resolution work [18] [19] with a similar assumption that for each pair of corresponding LRI and HRI blocks (patches), their local neighborhoods on some proper manifolds would be similar. More detailed, each HRI block $x_i$ and its nearest neighbors in high dimension lie on or close to a locally-linear manifold. This local structure can be characterized by a weighting vector $W_i$ which contains the linear coefficients that reconstruct $x_i$ from its $k$ nearest neighbors $\overline{x_j}$ selected from the high dimensional space. $W_i$ can be solved by minimizing the reconstruction error $\epsilon(W_i)$ in Eq.(6) with the constraint $\sum_{j=1}^{k} W_i(j) = 1$. Since low-resolution and high-resolution manifolds have similar structure, the weights $W_i$ minimizing $\epsilon(W_i)$ on the HRI blocks should also yield a small value when the data are replaced with the LRI blocks, and vice versa.

$$\epsilon(W_i) = (x_i - \sum_{j=1}^{k} W_i(j)\overline{x_j})^2 \tag{6}$$

Our AC coefficients inference model is also based on this assumption that HRI block set and LRI block set share the the similar locally-linear structure. But in our algorithm, this structure will be estimated more reasonably as shown in Eq.(11) by both considering the LRI block and HRI block manifolds instead of only considering one manifold.

In details, it is assumed that LRI blocks and HRI blocks differ from the approximations obtained with their $k$ nearest neighbors and weights in each corresponding manifold space by zero mean Gaussian noise of variance $\sigma_L^2$ and $\sigma_H^2$, respectively. Therefore, the two local geometry similarity can be described as:

$$I_L^{AC}(i) = \sum_{j=1}^{k} W_i(j)\overline{I_L^{AC}}(j) + N(0, \sigma_L^2) \tag{7}$$

$$I_H^{AC}(i) = \sum_{j=1}^{k} W_i(j)\overline{I_H^{AC}}(j) + N(0, \sigma_H^2) \tag{8}$$

Where $\overline{I_L^{AC}}(j)$ and $\overline{I_H^{AC}}(j)$ denote the nearest neighbor of $I_L^{AC}(j)$ and $I_H^{AC}(j)$ in training set $\Phi$, respectively. Unlike traditional learning-based work, this is a more general way to use the training data, because the target HRI block is generated depending on several nearest neighbors instead of only one. Based on Eq.(7) and Eq.(8), the compatibility function $\phi(I_H^{AC}(i), I_L^{AC}(i))$ can be defined as:

$$\phi(I_H^{AC}(i), I_L^{AC}(i)) = exp\{-(I_L^{AC}(i) - \sum_{j=1}^{k} W_i(j)\overline{I_L^{AC}}(j))^2$$
$$/2\sigma_L^2\} \times exp\{-(I_H^{AC}(i) - \sum_{j=1}^{k} W_i(j)\overline{I_H^{AC}}(j))^2/2\sigma_H^2\} \quad (9)$$

A factor $\lambda$ is introduced as $\lambda = \sigma_L^2/\sigma_H^2$, then

$$\phi(I_H^{AC}(i), I_L^{AC}(i)) = exp\{-((I_L^{AC}(i) - \sum_{j=1}^{k} W_i(j)\overline{I_L^{AC}}(j))^2$$
$$+ \lambda(I_H^{AC}(i) - \sum_{j=1}^{k} W_i(j)\overline{I_H^{AC}}(j))^2)/2\sigma_L^2\} \quad (10)$$

An energy term is introduced as

$$E(I_H^{AC}(i), W_i; I_L^{AC}(i)) = -2\sigma_L^2 \times ln(\phi(I_H^{AC}(i), I_L^{AC}(i)))$$
$$= E_1(I_L^{AC}(i), W_i) + \lambda E_2(I_H^{AC}(i), W_i) \quad (11)$$

Where

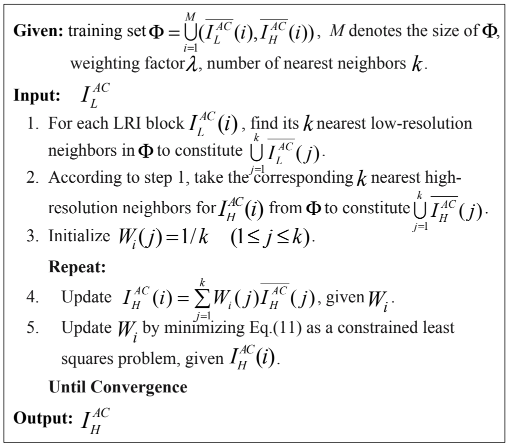
$$E_1(I_L^{AC}(i), W_i) = (I_L^{AC}(i) - \sum_{j=1}^{k} W_i(j)\overline{I_L^{AC}}(j))^2 \quad (12)$$

$$E_2(I_H^{AC}(i), W_i) = (I_H^{AC}(i) - \sum_{j=1}^{k} W_i(j)\overline{I_H^{AC}}(j))^2 \quad (13)$$

According to Eq.(10) and Eq.(11), the optimization of Eq.(5) can be solved by minimizing the energy $E(I_H^{AC}, W; I_L^{AC}) = \sum_i E(I_H^{AC}(i), W_i; I_L^{AC}(i))$.

### 3.2.2 Energy Minimization

The parameter $\lambda$ in Eq.(11) can be set empirically in the experiments because it effects as a weighting factor to balance the contributions of $E_1(I_L^{AC}(i), W_i)$ and $E_2(I_H^{AC}(i), W_i)$. Given the training set, the minimization of $E(I_H^{AC}, W; I_L^{AC})$ with respect to $I_H^{AC}(i)$ and to $W_i$ can be solved respectively and iteratively. For $ith$ block, firstly set $I_H^{AC}(i) = \sum_{j=1}^{k} W_i(j)\overline{I_H^{AC}}(j)$ with the initialized $W_i$. Then update $W_i$ with $I_H^{AC}(i)$ by minimizing $E(I_H^{AC}(i), W_i; I_L^{AC}(i))$ in Eq.(11) as a constrained least squares problem. The whole AC coefficients inference algorithm is summarized in Figure 8.



```
Given: training set $\Phi = \bigcup_{i=1}^{M}(\overline{I_L^{AC}}(i), \overline{I_H^{AC}}(i))$, $M$ denotes the size of $\Phi$,
        weighting factor $\lambda$, number of nearest neighbors $k$.
Input:  $I_L^{AC}$
  1. For each LRI block $I_L^{AC}(i)$, find its $k$ nearest low-resolution
     neighbors in $\Phi$ to constitute $\bigcup_{j=1}^{k}\overline{I_L^{AC}}(j)$.
  2. According to step 1, take the corresponding $k$ nearest high-
     resolution neighbors for $I_H^{AC}(i)$ from $\Phi$ to constitute $\bigcup_{j=1}^{k}\overline{I_H^{AC}}(j)$.
  3. Initialize $W_i(j)=1/k \quad (1 \le j \le k)$.
     Repeat:
  4.   Update $I_H^{AC}(i) = \sum_{j=1}^{k} W_i(j)\overline{I_H^{AC}}(j)$, given $W_i$.
  5.   Update $W_i$ by minimizing Eq.(11) as a constrained least
       squares problem, given $I_H^{AC}(i)$.
     Until Convergence
Output: $I_H^{AC}$
```

Figure 8. AC coefficients Inference Algorithm.

## 4. HRI Reconstruction by IDCT and Postfiltering

As shown in Figure 1, all DCT coefficients of each block in the target prefiltered HRI $I_H$ can be recovered by combining the following two parts: 1) the selected AC coefficients which constitute $I_H^{AC}(i)$ are estimated by the above AC coefficients inference model and other residue AC coefficients are set to zero; 2) interpolate the LRI to HRI by Cubic B-Spline method and the target DC coefficients are estimated from the corresponding block of the interpolated HRI. Given all DCT coefficients for each block, the target prefiltered HRI $I_H$ can be reconstructed by IDCT. Then the final HRI result $I_H^*$ can be derived from $I_H$ by the above postfiltering scheme.

## 5. Learning Block Dictionary by Clustering

As discussed in [5], the performance of learning-based method often depends on how well the input LRI matches the samples in the training set. Theoretically, the more training samples are collected, the more robust the learning-based algorithm is. However, a huge training set makes it difficult to design a fast algorithm due to the taxing computation and heavy memory load. Fortunately, the blocks cropped from the face images do not have much variation, since face images are similar and the subparts of face images are more similar. This is especially true in our case because our training set only contains AC coefficients which represent local facial features. The raw training set should have much redundancy and it is possible to learn those most representative blocks and build a compact block dictionary by performing clustering method.

In our training, all collected training images are firstly aligned by affine transform based on three marked points: the centers of the two eyes and the center of the mouth.

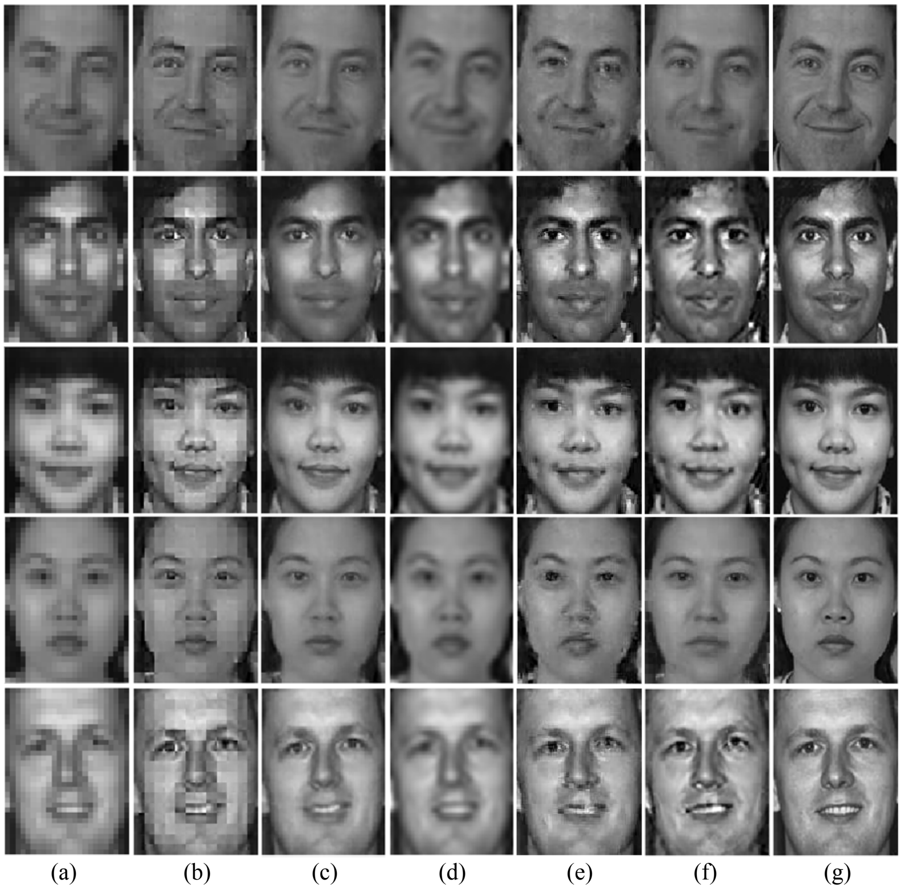


Figure 9. Face hallucination results and comparison. (a) input LRIs ($24 \times 32$); (b) our intermediate prefiltered results $I_H$; (c) our final results $I_H^*$; (d) Cubic B-Spline Interpolation; (e) Baker et al.; (f) C.Liu et al.; (g) original HRIs ($96 \times 128$).

Then each image is cropped to a canonical $96 \times 128$ image as the HRI. Its corresponding $24 \times 32$ LRI can be obtained by downsampling and smoothing. After being preprocessed by the above prefiltering scheme, all HRIs are transformed from spatial domain to frequency domain by $8 \times 8$ DCT. So the HRI blocks of the training data and testing data are non-overlapped $8 \times 8$ blocks and represented by only using the first 15 AC coefficients in zig-zag order. Since the LRIs will be initially enlarged via Cubic B-Spline interpolation, AC coefficients of the corresponding LRI blocks are obtained by performing $8 \times 8$ DCT similarly on the interpolated HRIs. Finally, the block dictionary is built by the clusters obtained by adopting the affinity propagation clustering method [20] on these raw training samples. Thus lots of redundancy of the raw training samples has been removed and a condensed training set is obtained. Besides, the dimension of block size is also reduced a lot as stated in Section 2. Therefore, our learning process is more efficient than traditional learning-based work.

# 6. Experimental Results

## 6.1. Comparison

The experiment is conducted with a large number of frontal face images from FERET data set [21] [22] and other collections, which consist of many different races, illuminations and types of face images. Among all these samples, about 1600 images are selected as training data and the remainder images are for testing.

Our approach is compared with some of the existing methods as shown in Figure 9. Cubic B-Spline interpolation suffers from severe blurring problem. Baker et al's method produces noisy results in some important facial features. C.Liu et al's results seem better and the whole visual quality is satisfying, but some subtle characteristics can not be generated correctly and smoothly, especially the details around eyes. It can be concluded that our method shows superiority over others on recovering smoothed HRI with high quality facial details.

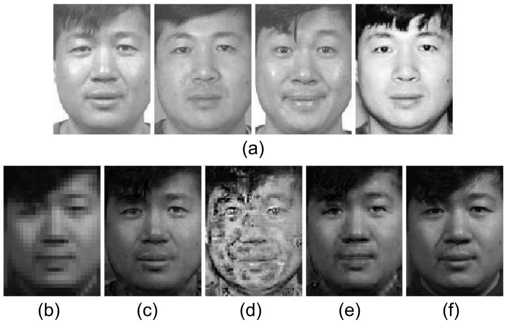


Figure 10. Face hallucination with a small training set. (a) training HRI priors; (b) input LRI (24 × 32); (c) our method; (d) learning in spatial domain; (d) Baker et al.; (f) original HRI (96 × 128).

### 6.2. Robustness to Image Illumination

As stated in Section 1 and 2, we only concern ourselves with learning local features embedded in AC coefficients from the training priors. It is found that without considering DC coefficients will make the learning process more robust since the matching from input to training samples is much less influenced by image illumination. An experiment as shown in Figure 10 is conducted to test the learning robustness of our method. All the five $96 \times 128$ images as shown in Figure 10 (a) and (f) are taken for the same person at different time, with different expressions and illumination conditions. Four images as shown in Figure 10 (a) with high illumination are selected for training, Figure 10 (f) captured under low illumination is used for testing. Given the LRI input Figure 10 (b), Figure 10 (c) is hallucinated by our method. It is obvious that our algorithm is nearly exempted from the illumination influence and capable of learning high quality local features from such a small training set. In contrast, Figure 10 (d) which is inferred by learning the pixel intensities directly (considering both DC and AC components) in spatial domain with an example-based manner, is very bad because the input LRI can not match well with the training samples due to the influence of image illumination. Although Baker et al's method produces a better result as shown in Figure 10 (e), it still fails in digging out some subtle features from the training samples. Since the training set is too small to be used in C.Liu et al's method, we exclude their method in this comparison.

## 7. Conclusion

In this paper, we presented an efficient learning-based framework for face hallucination from a single LRI. Our method is novel in that the problem is formulated as DCT coefficients estimation in frequency domain. DC coefficients were estimated by Cubic B-Spline interpolation. AC coefficients were learned efficiently from a condensed training block dictionary. Experiments clearly demonstrated the effectiveness and robustness of our approach.

## Acknowledgement

The authors would like to thank W. Liu, Y. Liu and all anonymous reviewers for their suggestions.

## 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. PAMI*, Vol. 24, No. 9, pp. 1167-1183, 2002.

[3] B. S. Morse and D. Schwartzwald. Image magnification using level-set reconstruction. In *Proc. of CVPR*, pp.333-340, 2001.

[4] Z. Lin and H. Y. Shum. Fundamental Limits of Reconstruction-Based Superresolution Algorithms under Local Translation. *IEEE Trans. PAMI*, Vol. 26, No. 1, pp. 83-97, 2004.

[5] Z. Lin, J. He, X. Tang and Chi-Keung Tang. Limits of Learning-Based Superresolution Algorithms. In *Proc. of ICCV*, 2007.

[6] W. T. Freeman, E. C. Pasztor and O. T. Carmichael. Learning Low-Level Vision. *International Journal of Computer Vision*, 40(1), pp.25-27, 2000.

[7] J. Sun, H. Tao and H. Y. Shum. Image Hallucination with Primal Sketch Priors. In *Proc. of CVPR*, pp. II 729-736, 2003.

[8] C. M. Bishop, A. Blake and B. Marthi. Super-resolution Enhancement of Video. In *Proc. Artificial Intelligence and Statistics*, 2003.

[9] C. Liu, H. Y. Shum and W. T. Freeman. Face hallucination: theory and practice. *International Journal of Computer Vision*, Vol. 75, No. 1, pp. 115-134, 2007.

[10] Q. Wang, X. Tang and H. Y. Shum. Patch Based Blind Image Super Resolution. In *Proc. of ICCV*, Vol. 1, pp. 709 - 716, 2005.

[11] W. Liu, D. Lin and X. Tang. Hallucinating Faces: TensorPatch Super-Resolution and Coupled Residue Compensation. In *Proc. of CVPR*, Vol. 2, pp. 478-484, 2005.

[12] Tuan. Q. Pham, L. J. van Vliet and K. Schutte. Resolution Enhancement of Low Quality Videos using a High-resolution Frame. In *Proc. of VCIP*, SPIE vol. 6077, 2006.

[13] Karl. S. Ni and T. Q. Nguyen. Image Superresolution Using Support Vector Regression. *IEEE Trans. on Image Processing*, Vol. 16 No. 6 pp. 1596-1610, 2007.

[14] T. D. Tran, J. Liang and C. Tu. Lapped transform via time-domain pre- and post-filtering. *IEEE Trans. on Signal Processing*, vol. 51, pp. 1557-1571, Jun. 2003.

[15] C. Tu and T. D. Tran. Context based entropy coding of block transform coefficients for image coding. *IEEE Trans. on Image Processing*, vol. 11, pp. 1277-1283, Nov. 2002.

[16] W. B. Pennebaker and J. L. Mitchell. *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.

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

[18] H. Chang, D. Y. Yeung and Y. Xiong. Super-Resolution through Neighbor Embedding. In *Proc. of CVPR*, pp. I: 275-282, 2004.

[19] Tien-Lung Chang, Tyng-Luh Liu and Jen-Hui Chuang. Direct Energy Minimization for Super-Resolution on Nonlinear Manifolds. In *Proc. of ECCV*, pp. 281-294, 2006.

[20] B. J. Frey and D. Dueck. Clustering by Passing Messages between Data Points. *Science*, Vol. 315, pp. 972-976, 2007.

[21] P. J. Phillips, H. Wechsler, J. Huang and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, Vol.16, No.5, pp. 295-306, 1998.

[22] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss. The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans. PAMI*, Vol. 22, pp. 1090-1104, 2000.