A Two-Step Approach to Hallucinating Faces:
Global Parametric Model and Local Nonparametric Model

Ce Liu*[†]   Heung-Yeung Shum[†]   Chang-Shui Zhang*

*State Key Lab of Intelligent Technology and Systems, Dept. of Automation, Tsinghua University, Beijing 100084, China
[†]Visual Computing Group, Microsoft Research China, Sigma Building, Zhi-Chun Road, Beijing 100080, China

Abstract

In this paper, we study face hallucination or synthesizing a high-resolution face image from a low-resolution input, with the help of a large collection of other high-resolution face images. We develop a two-step statistical modeling approach that integrates both a global parametric model and a local nonparametric model. First, we derive a global linear model to learn the relationship between the high-resolution face images and their smoothed and down-sampled lower resolution ones. Second, the residual between an original high-resolution image and the reconstructed high-resolution image by learned linear model is modeled by a patch-based nonparametric Markov network, to capture the high-frequency content of faces. By integrating both global and local models, we can generate photorealistic face images. Our approach is demonstrated by extensive experiments with high-quality hallucinated faces.

1. Introduction

Super-resolution techniques in computer vision infer the missing high-resolution image from the low-resolution input. Low-resolution is equivalent to low-frequency and high-resolution consists of high, middle and low frequency bands. There are in general two classes of super-resolution techniques: reconstruction-based (from input images alone) and learning-based (from other images). Of particular interest is *face hallucination*, or learning high-resolution face images from low-resolution ones. Face hallucination is a term coined by Baker and Kanade [1], which implies the high-frequency part of face image must be purely fabricated. Hallucinating faces is particularly challenging because people are so familiar with faces.

We argue that a successful face hallucination algorithm should meet the following three constraints:

1) **Sanity constraint.** The result must be very close to the input image when smoothed and down-sampled.
2) **Global constraint.** The result must have common characteristics of a human face, *e.g.* eyes, mouth and nose, symmetry, etc.
3) **Local constraint.** The result must have specific characteristics of this face image with photorealistic local features.
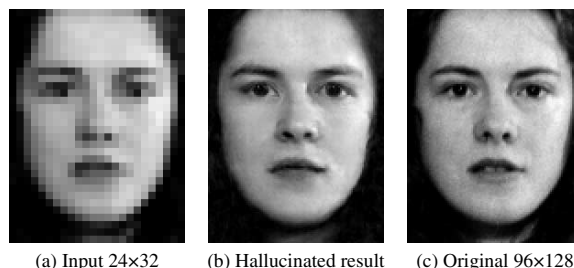
(a) Input 24×32   (b) Hallucinated result   (c) Original 96×128

Figure 1. An example of face hallucination using our approach. The hallucinated image (b) is locally different from, but globally similar to the original high-resolution one (c).

The first requirement can be easily made. For example, it can be simply formulated as a linear constraint on the result. The second and third ones are apparently much harder. It is important to satisfy the three requirements altogether to hallucinate faces with good quality. Without the constraint on specific face features, the result would be too smooth, close to the average face. On the other hand, without a global face similarity constraint, the result could be noisy.

Such global and local constraints lead us to a hybrid approach. We combine a global parametric model which generalizes well with common faces, with a local nonparametric model which learns local textures from example faces. This approach can also be applied to modeling other visual patterns, in particular for the structural objects with both global coherence such as illumination, contour and symmetry, and precise local textures like skin and hair.

We incorporate all the constraints in a statistical face model and find the maximum *a posteriori* (MAP) solution for the hallucinated face. An example of a hallucinated image from an input low-resolution image with our approach is shown in Figure 1. The sanity constraint is simply modeled as a Gaussian distribution or a soft constraint. The global constraint assumes a Gaussian distribution learnt by principal component analysis (PCA). The local constraint utilizes a patch-based nonparametric Markov network to learn the statistical relationship between the global face image and the local features. A two-step approach is then used in hallucinating faces. First, an optimal global face image is obtained in the eigen-space when constraints 1) and 2) are satisfied. Second, an optimal local feature image is inferred

from the optimal global image by minimizing the energy of the Markov network with constraint 3) applied. The sum of the global and local images forms the final result.

After reviewing related work in Section 2, we introduce the details of our model in Section 3. Many convincing examples are shown in Section 4. Section 5 gives discussion and conclusion.

2. Related Work

Most learning-based super-resolution algorithms such as [3, 6, 7] assume homogeneous (stationary) Markov random fields (MRFs) for images. Let L denote an image lattice, and $\mathbf{v}$ a certain position on the lattice with $I_{\mathbf{v}}$ as the pixel value. $I_{\mathbf{v}}^-$ represents all pixels on L other than $I_{\mathbf{v}}$. L is a Markov random field if

$$p(I_{\mathbf{v}}|I_{\mathbf{v}}^-) = p(I_{\mathbf{v}}|N_{\mathbf{v}}), \tag{1}$$

where $N_{\mathbf{v}}$ is the neighborhood of $\mathbf{v}$ [14]. This definition indicates that a pixel only relies on the pixels in its neighborhood. Further, L is a *homogeneous* MRF if for any positions $\mathbf{v}$ and $\mathbf{u} \in L$, their conditional density functions are identical, *i.e.*,

$$p(I_{\mathbf{v}}|N_{\mathbf{v}}) = p(I_{\mathbf{u}}|N_{\mathbf{u}}), \tag{2}$$

which also implies that their neighborhoods have equal size.

Proposed for texture synthesis, the multi-resolution non-parametric sampling method developed by De Bonet [3] indeed infers the high-frequency features from the low-frequency features named the *parent structure*. It is demonstrated that in a homogeneous MRF, the high-frequency component locally depends on the low-frequency part. Freeman and Pasztor [6] proposed a parametric Markov network to learn the statistics between the "scene" and "image", as a framework for handling low-level vision tasks. It can be applied in super resolution work if the scene and image are high- and low-frequency bands respectively. Recently Hertzmann *et al.* [7] generalizes local feature transform methods in "Image Analogies". Given a pair of training images, an analogous image is inferred from the input by the local similarity between the training pair. It can also fulfill super resolution objectives if the training pairs are high- and low-resolution images, respectively.

All of above methods do local feature transfer/inference with low-level vision. They perform well in hallucinating texture-like images, but would fail in hallucinating structural visual patterns such as human faces. To broaden the application to face hallucination, the homogeneous MRF assumption (2) has to be abandoned, leading to the work by Baker and Kanade [1]. They only follow the MRF assumption in that the size of each pixel's neighborhood is equal. The statistics between the low- and high-resolution images at each position is learnt in a nonparametric way by a number of training pairs. Similar to [3], the features on a high-frequency image are inferred from the parent structure by

nearest neighbor searching. They also discuss the limits on super resolution and how to break them in their method [2]. The results in [1] appear to be noisy at places. Moreover, some global properties of a face, such as explicit contours, coherent illumination and symmetry are also missed.

It is interesting to note that all previous models use local feature inference or transfer in MRF without global correspondence being taken into account. Such global modeling is, however, essential for achieving good performance in face hallucination. Principal component analysis (PCA) has been successfully used in face recognition by eigenfaces [12] and face modeling by ASM and AAM [4]. We use PCA to model the global variance of facial appearance. Since patch-based nonparametric sampling has been applied in texture synthesis with encouraging results and high efficiency [8, 13, 5], we build a patch-based Markov network, a nonparametric counterpart to the parametric one proposed in [6], to model the statistics between a local feature image and global face image. The theory and algorithm are explained in the next section.

3. Theory and Algorithms

3.1 A Bayesian Formulation to Face Hallucination

Let I_H and I_L denote the high- and low-resolution face images respectively. We simply choose the same method as that in [1] to compute I_L from I_H. If I_L is s times smaller than I_H, I_L is computed by

$$I_L(m,n) = \frac{1}{s^2} \sum_{i=0}^{s-1} \sum_{j=0}^{s-1} I_H(sm+i, sn+j) \tag{3}$$

where s is always an integer with the default value 4 in this article. Equation (3) combines a smoothing step and a down-sampling step, more consistent with image formation as integration over the pixel [1]. To simplify the notation, if I_H and I_L are respectively N-D and M-D long vectors ($M = N/s^2$), equation (3) can be rewritten as

$$I_L = A I_H \tag{4}$$

where $A = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_M]^T$ is a $M \times N$ matrix with each row vector $\mathbf{a}_i^T$ smoothing a $s \times s$ block in I_H to a pixel in I_L.

To compute I_H from I_L is straightforward in (4), but the inverse process is full of uncertainty, with uncountable I_Hs (almost) satisfying (4). Based on the maximum *a posteriori* (MAP) criterion, we find the optimal solution maximizing the posterior probability $p(I_H|I_L)$, *i.e.*,

$$I_H^* = \arg \max_{I_H} p(I_L|I_H)p(I_H). \tag{5}$$

Under this framework, we should (a) build both the prior and likelihood models and (b) find the optimal solution.

3.2 Global and Local Modeling of Face

Note that in equation (5) exists the prior distribution of the face image $p(I_H)$. Current face prior models either capture common properties of face such as eigenfaces [12] and

AAM [4], or represent individual characteristics such as local features [1]. But both of them are required in face hallucination. We develop a mixture model combining a global parametric model called *global face image* I_H^g carrying common facial properties, and a local nonparametric one called *local feature image* I_H^l recording local individualities, illustrated in Figure 2. The high-resolution face image is naturally a composition of them,

$$I_H = I_H^l + I_H^g. \tag{6}$$

Since I_L is the low-frequency part of I_H, the global face I_H^g contributes the main part of AI_H and the local feature I_H^l lies on the high-frequency band. Mathematically,

$$AI_H^g = AI_H, \ AI_H^l = 0. \tag{7}$$

In this way, the prior model of face is decomposed to

$$p(I_H) = p(I_H^l, I_H^g) = p(I_H^l|I_H^g)p(I_H^g). \tag{8}$$

Now we shall reformulate the MAP problem (5) under this mixture face model. The likelihood $p(I_L|I_H)$ is simply regarded as a soft constraint to I_H, and exhibits a Gaussian form if each pixel on I_L is identically treated [1]

$$p(I_L|I_H) = \frac{1}{Z} \exp\{-\|AI_H - I_L\|^2/\lambda\}, \tag{9}$$

where Z is a normalization constant, λ scales the variance and $\|x\|^2 = x^Tx$ throughout this paper. Based on (7), (9) can be rewritten as

$$p(I_L|I_H) = \frac{1}{Z} \exp\{-\|AI_H^g - I_L\|^2/\lambda\} = p(I_L|I_H^g). \tag{10}$$

From equation (8) and (10), the MAP inference problem (5) can be transferred to

$$I_H^* = \arg \max_{I_H^g, I_H^l} p(I_L|I_H^g)p(I_H^g)p(I_H^l|I_H^g). \tag{11}$$

Obviously $p(I_L|I_H^g)p(I_H^g)$ and $p(I_H^l|I_H^g)$ sequentially constrain I_H^g and I_H^l. The solution strategy is naturally divided into two steps. In the first step we leave I_H^l apart and obtain the optimal global face I_H^{g*} by maximizing $p(I_L|I_H^g)p(I_H^g)$. In the second stage the optimal local feature image I_H^{l*} is computed by maximizing $p(I_H^l|I_H^g)$. Finally $I_H^* = I_H^{g*} + I_H^{l*}$ is our desired result.

3.3 Global Modeling: A Linear Parametric Model

We apply PCA to modeling the global face image I_H^g. Given a set of training face images $\{I_H^{(i)}\}_{i=1}^k$, we can compute the eigenvectors $\{\mathbf{b}_i\}_{i=1}^l (\mathbf{b_i} \in \mathbf{R}^N)$, eigenvalues $\{\sigma_i^2\}_{i=1}^l$ and mean face μ with l the reduced dimension. The orthogonal eigenvectors construct the eigen-subspace $\Omega = span(\mathbf{b}_1, \cdots, \mathbf{b}_l) \sim \mathbf{R}^l$. Thus I_H^g is in fact the reconstructed image of I_H in Ω

$$I_H^g = BX + \mu, X = B^T(I_H - \mu), \tag{12}$$

where $B = [\mathbf{b}_1, \cdots, \mathbf{b}_l]_{N \times l}$, and $X = (x_1, \cdots, x_l)^T$ is a vector in Ω. Intuitively I_H^g is linearly controlled by x_is with corresponding eigenvectors $\mathbf{b}_i$s. Since the eigenvectors are analyzed from the training data representing the irrelevant common face properties such as lighting, scale and pose, I_H^g retains the common features of I_H with individuality lost.

The distribution of I_H^g can be purely replaced by X based on (12). Maximizing $p(I_L|I_H^g)p(I_H^g)$ in (11) is equivalent to maximizing $p(I_L|X)p(X)$. The prior $p(X)$ is simply assumed as Gaussian:

$$p(X) = \frac{1}{Z'} \exp\{-X^T\Lambda^{-1}X\}, \tag{13}$$

where $\Lambda = \text{diag}(\sigma_1^2, \cdots, \sigma_l^2)$ and Z' is a normalization constant. The likelihood (10) is replaced by

$$p(I_L|X) = \frac{1}{Z} \exp\{-\|A(BX+\mu) - I_L\|^2/\lambda\}. \tag{14}$$

To maximize $p(I_L|X)p(X)$ is

$$X^* = \arg \min_X \lambda X^T\Lambda^{-1}X + \|A(BX+\mu) - I_L\|^2, \tag{15}$$

with solution:

$$X^* = (B^TA^TAB + \lambda\Lambda^{-1})^{-1}B^TA^T(I_L - A\mu). \tag{16}$$

The optimal global face image $I_H^{g*} = BX^* + \mu$. Since matrix B, Λ and μ are learnt by PCA, and A is constant as a smoothing and down-sampling function, all matrices on the right side of (16) can be off-line computed and I_H^{g*} is calculated very fast input I_L. I_H^{g*} is very close to human face with some smoothness, which will be improved by the local model in next subsection.

3.4 Local Modeling: Patch-based Nonparametric Markov Network

In most cases PCA is used, the random variable is regarded as a composition of two parts: principal components and independent residual. But in our mixture modeling, the residual $I_H^l = I_H - I_H^g$ is the highest frequency component, relying on rather than independent to the lower frequency part I_H^g [6]. To carefully model $p(I_H^l|I_H^g)$, we combine Markov network [6] and "Image Analogies" [7] to generate a new patch-based nonparametric Markov network, in which the statistics between two images are modelled by connected patches in a nonparametric way. The function of the Markov network is shown in Figure 2.

Let I_H^l and I_H^h be decomposed to square patches with size $w+h$. $I_H^l(m,n)$ and $I_H^g(m,n)$ denote patches centered at position $((m+\frac{1}{2})w, (n+\frac{1}{2})w)$ respectively. For simplification, we replace (m,n) by a vector $\vec{v}$. For patch $I_H^l(\vec{v})$, its neighboring patches are $N_H^l(\vec{v}) = \{I_H^l(\vec{v} + \Delta_x), I_H^l(\vec{v} - \Delta_x), I_H^l(\vec{v} + \Delta_y), I_H^l(\vec{v} - \Delta_y)\}$ with overlapping size h and the overlapping area $S(\vec{v}) = I_H^l(\vec{v}) \cap N_H^l(\vec{v})$, where $\Delta_x = (1,0), \Delta_y = (0,1)$. Such circumstance varies slightly at the image boundary. I_H^l can be reconstructed by superposing
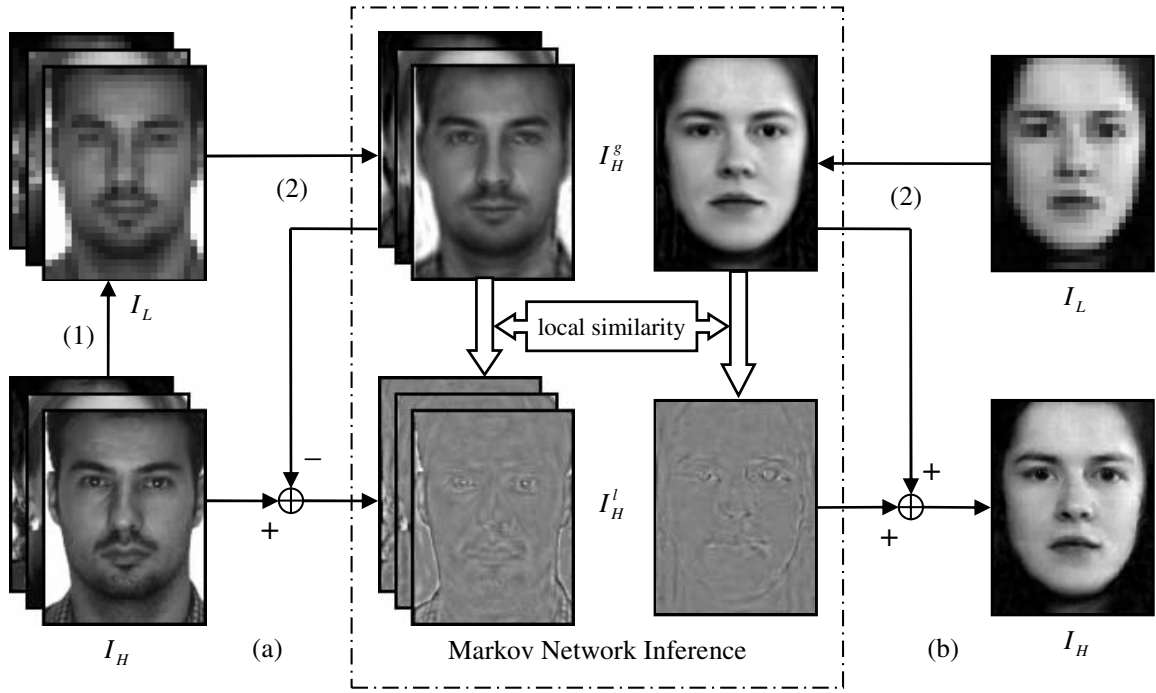
Figure 2. The function of Markov network in our model. (a) is the training process and (b) the hallucinating process. (1): smooth and down-sampling. (2): MAP inference to get the optimal global face I_H^{g*}. The Markov network finds the optimal local feature image I_H^{l*} by energy minimization.

its patches, with pixels in the overlapping area blended. The patches at the same position, $I_H^l(\vec{v})$ and $I_H^g(\vec{v})$, are connected between I_H^l and I_H^g. Thus all patches in these two images constitute a network, as illustrated in Figure 3.

Suppose $I_H^{l-}(\vec{v})$ denotes all patches on I_H^l except patch $I_H^l(\vec{v})$. We assume that the above network is a Markov network by defining:

$$p(I_H^l(\vec{v})|I_H^{l-}(\vec{v}), I_H^g) = p(I_H^l(\vec{v})|N_H^l(\vec{v}), I_H^g(\vec{v})), \quad (17)$$

This definition is equivalent to that in [6], but not homogeneous because the conditional density function of each patch is not identical. Suppose $p(I_H^l(\vec{v})|N_H^l(\vec{v}), I_H^g(\vec{v}))$ a Gibbs distribution

$$p(I_H^l(\vec{v})|N_H^l(\vec{v}), I_H^g(\vec{v})) \propto \exp\{-E_G(I_H^l(\vec{v}), N_H^l(\vec{v}), I_H^g(\vec{v}))\} \quad (18)$$

where $E_G(\cdot)$ is the Gibbs potential function to describe how likely a patch $I_H^l(\vec{v})$ connects to $I_H^g(\vec{v})$ and is surrounded by $N_H^l(\vec{v})$. It is natural to decouple it into two terms concerning $N_H^l(\vec{v})$ and $I_H^g(\vec{v})$ independently,

$$
\begin{aligned}
&E_G(I_H^l(\vec{v}), N_H^l(\vec{v}), I_H^g(\vec{v}))\\
&= E_G^{int}(I_H^l(\vec{v}), N_H^l(\vec{v})) + E_G^{ext}(I_H^l(\vec{v}), I_H^g(\vec{v})) \quad (19)\\
&\equiv E_G^{int}(\vec{v}) + E_G^{ext}(\vec{v})
\end{aligned}
$$

where $E_G^{int}(\vec{v})$ is the *internal* potential function that describes the neighboring statistics between patches inside I_H^l, and $E_G^{ext}(\vec{v})$ is the *external* potential function that represents the connecting statistics between connecting patches in I_H^l and I_H^g.
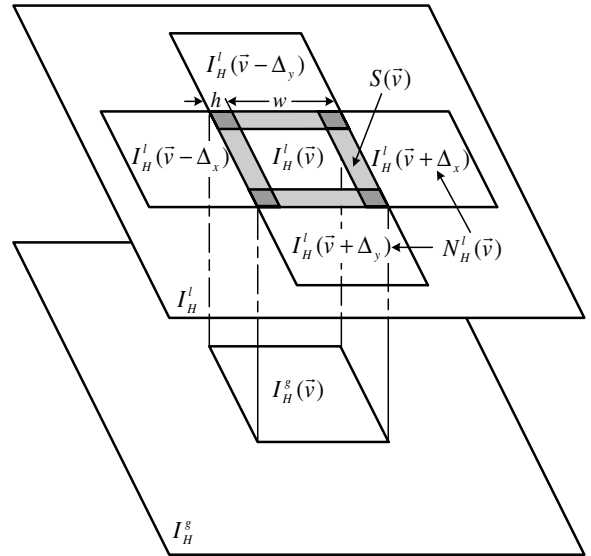

Figure 3. Illustration of the patch-based Markov network.

The *external* potential function $E_G^{ext}(\vec{v})$ is modeled upon training examples. Suppose at position $\vec{v}$ we have k training pairs $\{I_H^{l(i)}(\vec{v}), I_H^{g(i)}(\vec{v})\}_{i=1}^k$, then

$$E_G^{ext}(\vec{v}) = \frac{1}{\lambda'} \sum_{i=1}^k \delta[I_H^l(\vec{v}) - I_H^{l(i)}(\vec{v})] d^2[I_H^g(\vec{v}), I_H^{g(i)}(\vec{v})], \quad (20)$$

where $\delta(\cdot)$ is the dirac function, $d(\cdot)$ is the distance metric between two patches in I_H^g, and λ' scales the variance. The distance plays a crucial role in nonparametric models, and can be chosen as SSD (sum of squared differences) on

pixel value or on feature images such as pyramid and parent structures [3, 1, 7, 8]. Since the signals in the high-frequency band depend on those in the lower frequency band [6], here we define the distance on the Laplacian image L_H^g of I_H^g, which in fact represents the middle-frequency band of the face image. The squared distance between $I_H^g(\vec{v})$ and $I_H^{g(i)}(\vec{v})$ is

$$d^2[I_H^g(\vec{v}), I_H^{g(i)}(\vec{v})] = \|L_H^g(\vec{v}) - L_H^{g(i)}(\vec{v})\|^2. \qquad (21)$$

Equation (20) and (21) form a nonparametric distribution: compare the given patch $I_H^g(\vec{v})$ with training patches $\{I_H^{g(i)}(\vec{v})\}_{i=1}^k$, then the patch $I_H^{l(i)}(\vec{v})$ with $I_H^{g(i)}(\vec{v})$ close to $I_H^g(\vec{v})$ is most probable to be chosen as $I_H^l(\vec{v})$. This is the key principle of modeling example-based conditional density [3].

The *internal* potential function $E_G^{int}(\vec{v})$ is introduced to make neighboring patches well connected. Since the neighboring patches overlap each other, it enforces the common part of the abutting patches to be as similar as possible in the overlapping area. Mathematically it is defined by:

$$E_G^{int}(\vec{v}) = \frac{1}{\lambda''} \sum_{\mathbf{u} \in S(\vec{v})} [\, I_H^l(\mathbf{u}) - N_H^l(\mathbf{u}) \,]^2, \qquad (22)$$

where $I_H^l(\mathbf{u})$ is the pixel value of I_H^l and λ'' scales the variance. The total energy E_{MN} of the Markov network is the sum of each patch's energy:

$$E_{MN} = \sum_{\vec{v}} (E_G^{int}(\vec{v}) + E_G^{ext}(\vec{v})). \qquad (23)$$

We have mentioned in 3.2 that after the optimal global face I_H^{g*} is obtained, the optimal local feature I_H^{l*} is found to maximize $p(I_H^l | I_H^{g*})$. This is equivalent to minimizing the total energy E_{MN} of the Markov network, *i.e.*

$$I_H^{l*} = \arg \min_{I_H^l} E_{MN}. \qquad (24)$$

Finding the global minimum of E_{MN} is not trivial. If there are totally $R \times C$ patches in I_H^l and k pairs of training images, the solution space is as huge as $k^{R \times C}$! A greedy algorithm which sequentially finds $I_H^{l*}(\vec{v})$ by minimizing $E_G(\vec{v})$ such as [7], can only find a local optimum. We use *simulated annealing* which has been successfully used in many areas to find a global or satisfactory optimum. In every step when we need to flip $I_H^l(\vec{v})$, each candidate patch in $\{I_H^{l(i)}(\vec{v})\}_{i=1}^k$ is selected as $I_H^l(\vec{v})$ with probability proportional to

$$exp\{-(E_G^{int}(\vec{v}) + E_G^{ext}(\vec{v}))/T\}, \qquad (25)$$

where T is the temperature. In each loop, every patch in image I_H^l is sequentially randomly flipped. By gradually decreasing T to zero, I_H^l will converge to the global optimal I_H^{l*}. The pseudo code is listed below.

MinimizeMarkovNetworkEnergy

Loop until $T < \epsilon$
  Loop $\vec{v}$ sequentially visiting all patches
    compute the energy of each patch $I_H^{l(i)}(\vec{v})$
    set $I_H^l(\vec{v}) = I_H^{l(i)}(\vec{v})$ with probability (25)
  decrease T
$I_H^{l*} = I_H^l$

---

4. Experimental Results

Our experiments are conducted with a large number of frontal face images in the FERET data set [10], AR data set [9] and other collections, involving all kinds of races, illuminations and types of face. We select 1114 images as the training data, and other 39 images as the test data. Before experiment we manually align the face image with five points: the centers of the eyeballs, the tip of the nose and the corners of the mouth. The centers of the eyeballs are used to calibrate the orientation, and the five together are used to calibrate the scale. We do not choose the affine warp to retain the shape of a face. After a similarity transform, each image is aligned to a canonical 96×128 pixel image. The high-resolution image is smoothed and down-sampled to a low-resolution 24×32 image by (3).

We use standard SVD [11] to do principal component analysis of the training images. The dimension of the face images is reduced to 200 to retain 97% of the eigenvalue total. Since the number of the training data is much larger than the reduced dimension, the inverse of the covariance-like matrix in equation (16) does exist. In solution (16), λ controls the balance between the similarity of I_H^g to input I_L and that to mean face μ. We find values in $0.05 \sim 0.2$ appropriate for λ and we choose 0.1 in the experiment. Some obtained I_H^{g*}s are displayed in Figure 4(b). Compared with the low-resolution input, such solutions have more explicit contours and edges, because the principal components record the common variance of face image.

Further we apply the patch-based nonparametric Markov network to infer the optimal local feature image. We choose $w = 5$, $h = 2$ as the patch size and overlapping size. In fact, the results vary little if w is chosen between 4 and 6. The parameters λ' and λ'' control the tradeoff between the *external* and *internal* potentials. We simply choose $\lambda' = \lambda'' = 3wh$. The initial value of the temperature T is set to 1. In the simulated annealing process, T gradually decreases 10% at each time step. Finally we get the hallucinated face I_H^* by summing I_H^{l*} and I_H^{g*} up, shown in Figure 4(c). We may see that these results suit the three goals of face hallucination very well, visually very close to the original high images, Figure 4(d). Note that each part of the hallucinated face such as the eyes and nose, are different from that in the original image but as detailed as that in the real face image. The runtime of computing the global face is less than 1 second
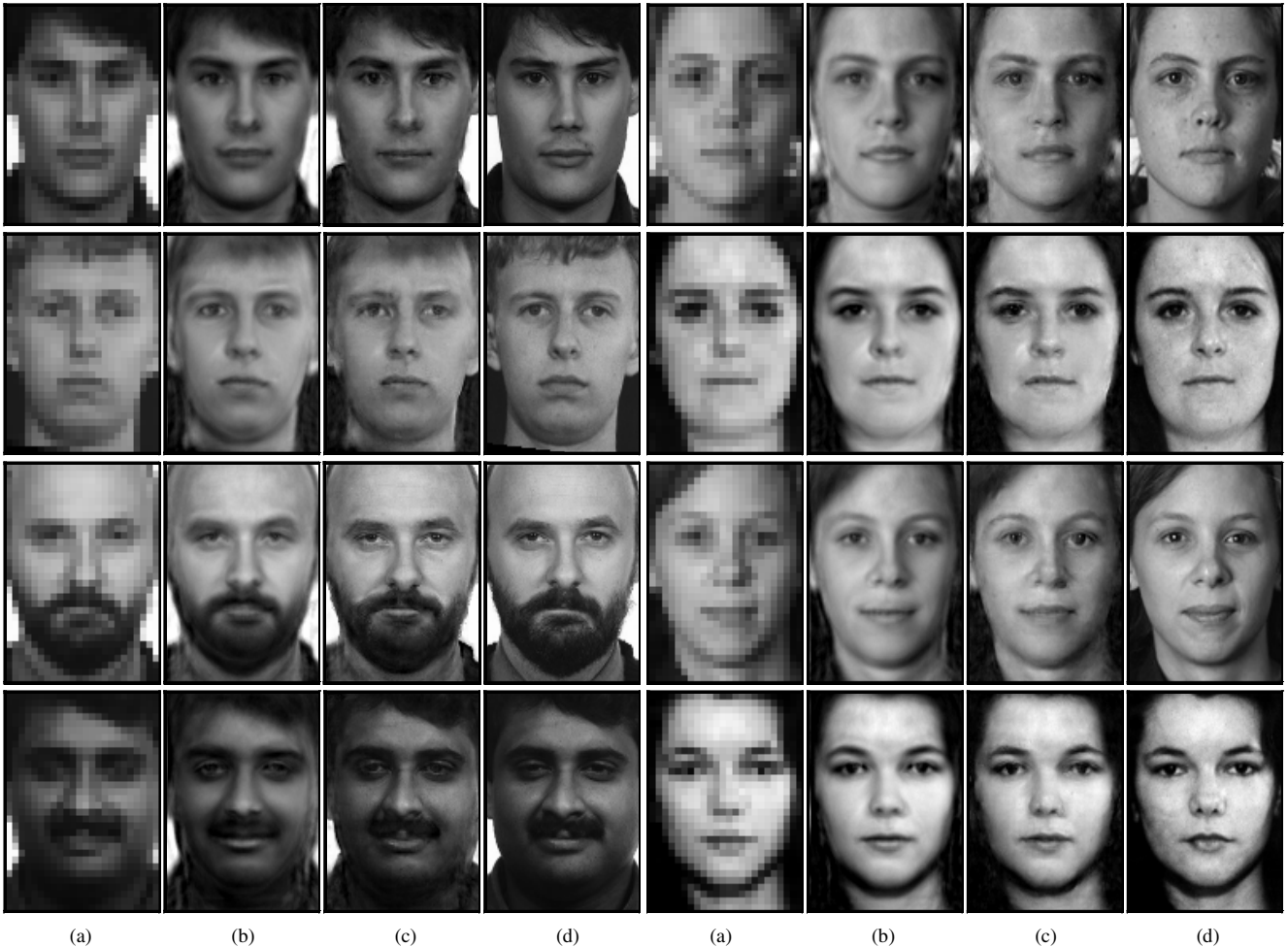
| (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |

Figure 4. The hallucination results. (a) is the low-resolution 24×32 input. (b) is the inferred global face I_H^{g*} from (a). (c) is the final result $I_H^* = I_H^{g*} + I_H^{l*}$. I_H^{l*} is inferred from I_H^{g*} by Markov network. (d) is the original high-resolution 96×128 image.

and the simulated annealing takes about half a minute.

We compare our algorithm with existing methods in Figure 5 for another group of individuals, using the same training data set. Those methods include (c) Cubic B-Spline , (d) Hertzmann's "Image Analogy" (We implement it in a patch-based way, also a nonparametric counterpart of Freeman's Markov network.) and (e) Baker's method. (d) is locally close to the face image but lacks global face features such as symmetry. (e) has too much noise also with global features missed. It is obvious that the result of our model (b) with both the global and local face information taken into account, is closer to the original face (f) with very high image qualities.

5. Discussion and Conclusion

Under MAP criterion, the kernel of face hallucination problem is how to model the low-resolution constraint and the appearance of face. A parametric model is proper to capture the common structural properties whereas a non-parametric model is appropriate to represent local and in-

dividual patterns. It is natural for us to combine them in face hallucination to integrate their merits. A linear model is used to model the global variance of face and the low-resolution constraint to achieve robustness and efficiency. We do not choose more complex models as the global face model because the modeling error can be compensated in the local model. We devise a patch-based nonparametric Markov network, a combination of [6, 7], to learn the relationship between local facial feature and global face. Non-parametric ensures it accuracy and patch-based endows it with high efficiency. The simulated annealing method here is very similar to the Gibbs sampling in [14], but converges more quickly. This method can also be expanded to general super-resolution problem.

To hallucinate high-quality face images, we develop a two-step statistical inference model which integrates both a global parametric linear model and a local nonparametric one. The effectiveness of our approach is demonstrated by extensive experiments.

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

Figure 5. Comparison between our method and others.

References

[1] S. Baker and T. Kanade. Hallucinating faces. *Fourth International Conference on Automatic Face and Gesture Recognition*, March 2000.

[2] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE conference on Computer Vision and Pattern Recognition*, June 2000.

[3] J. D. Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. *Proceedings of SIGGRAPH 97*, pages 361–368, August 1997.

[4] T. Cootes and C. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, 2000.

[5] A. Efros and W. Freeman. Quilting for texture synthesis and transfer. *Proceedings of SIGGRAPH 2001*, August 2001.

[6] W. Freeman and E. Pasztor. Learning low-level vision. *7-th International Conference on Computer Vision*, pages 1182–1189, 1999.

[7] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin. Image analogies. *Proceedings of SIGGRAPH 2001*, August 2001.

[8] L. Liang, C. Liu, Y. Xu, B. Guo, and H. Shum. Real-time texture synthesis by patch-based sampling. *ACM Transaction on Graphics*, 2001.

[9] A. Martinez and R. Benavente. The ar face database. Technical report, CVC Techinical Report No.24, June 1998.

[10] P. Philips, H. Moon, P. Pauss, and S. Rivzvi. The feret evaluation methodology for face-recognition algorithms. *In Proceedings of CVPR'97*, pages 137–143, 1997.

[11] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.

[12] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neurosciences*, 3:71–86, 1991.

[13] Y. Xu, B. Guo, and H. Shum. Chaos mosaic: fast and memory efficient texture synthesis. Technical report, In Microsoft Research Technical Report MSR-TR-2000-32, April 2000.

[14] S. Zhu, Y. Wu, and D. Mumford. Minimax entropy principle and its application to texture modeling. *Neural Computation*, 9(8), November 1997.