CHAPTER III

THE PROPOSED FRAMEWORKS

In this chapter, the two frameworks were proposed for improving the performance of face hallucination. In Section 3.1, the first framework, *Color Face Hallucination with Linear Regression Model in MPCA*, take the advantage of the MPCA. The second framework, *Color Face Super-Resolution Based on Tensor Patches Method*, was introduced in Section 3.2.

3.1 Color Face Hallucination with Linear Regression Model in MPCA

The goal of this part is to propose a novel hallucination reconstruction, using the MPCA for color face image. Given a training image set $\{\mathcal{X}^H, \mathcal{X}^L\}$, where $\mathcal{X}^H = \{\mathcal{X}_i^h\}_{i=1,\ldots,K}$, $\mathcal{X}^L = \{\mathcal{X}_i^l\}_{i=1,\ldots,K}$ and K is the number of training image. The training color face images can be defined as $\mathcal{X}_i^h \in \mathbf{R}^{I_1} \otimes \mathbf{R}^{I_2} \otimes \mathbf{R}^{I_3}$ and $\mathcal{X}_i^l \in \mathbf{R}^{J_1} \otimes \mathbf{R}^{J_2} \otimes \mathbf{R}^{J_3}$ which are the HR color face image and LR color face image, respectively.

In this paper, we propose our method in many color models such as RGB, YCbCr, HSV and CIELAB. For example, the training image sets $\mathcal{X}_i^l$ and $\mathcal{X}_i^h$ can be described in RGB model as $\mathcal{X}_i^h \in \mathbf{R}^{I_1 \times I_2 \times 3}$ and $\mathcal{X}_i^l \in \mathbf{R}^{J_1 \times J_2 \times 3}$. In addition, the index $I_1 \times I_2 \times 3$ is an array of color pixels, where each color pixel is a triplet corresponding to the red, green, and blue components of an RGB image.

Following standard multilinear algebra, any tensor can be expressed as the product

$$\mathcal{Y}_i^h = \mathcal{X}_i^h \times_1 \widetilde{\mathbf{U}}^{h(1)^T} \times_2 \widetilde{\mathbf{U}}^{h(2)^T} \times_3 \widetilde{\mathbf{U}}^{h(3)^T} \tag{3.1}$$

and

$$\mathcal{Y}_i^l = \mathcal{X}_i^l \times_1 \widetilde{\mathbf{U}}^{l(1)^T} \times_2 \widetilde{\mathbf{U}}^{l(2)^T} \times_3 \widetilde{\mathbf{U}}^{l(3)^T} \tag{3.2}$$

where $\mathcal{Y}_i^h \in \mathbf{R}^{P_1} \otimes \mathbf{R}^{P_2} \otimes \mathbf{R}^{P_3}$ and $\mathcal{Y}_i^l \in \mathbf{R}^{Q_1} \otimes \mathbf{R}^{Q_2} \otimes \mathbf{R}^{Q_3}$, with ($P_1 < I_1, P_2 < I_2$ and $P_3 < 3$) as the index of HR training set and ($Q_1 < J_1, Q_2 < J_2$ and $Q_3 < 3$) as the index of LR training set. The tensor $\mathcal{Y}_i$ can capture most of the variations observed in the original tensor objects, assuming that these variations are measured by the total scatter. Therefore, two sets of MPCA subspace projection are obtained, which are $\mathcal{Y}_i^h = [y_{r,s,t}^h]_i$ and $\mathcal{Y}_i^l = [y_{r,s,t}^l]_i$ respectively. In addition, we use $[y_{r,s,t}]$ to represent a tensor with $y_{r,s,t}$ as its (r, s, t)-th entry.

One can see that if the sets of $\widetilde{\mathbf{U}}^{(1)}, \widetilde{\mathbf{U}}^{(2)}, \widetilde{\mathbf{U}}^{(3)}$ are disjoint sets of orthonormal vectors then the correlation between the decomposition coefficients can be suppressed. From the model $\mathcal{Y}_i^h = f(\mathcal{Y}_i^l)$. When a generative model is used, f is actually a probability. Thus we can consider the conditional probability $P(\mathcal{Y}_i^h|\mathcal{Y}_i^l)$. When a new testing color face image

(LR) $\mathcal{X}^l$ is provided, the HR MPCA subspace projection is given by:

$$\mathcal{Y}^h = \underset{\mathcal{Y}}{\operatorname{argmax}} P(\mathcal{Y}|\mathcal{Y}^l). \tag{3.3}$$

The HR color face image $\mathcal{X}^h$ can be reconstructed by back-projection from the MPCA subspace into the image tensor space as

$$\mathcal{X}^h = \mathcal{Y}^h \times_1 \widetilde{\mathbf{U}}^{h(1)} \times_2 \widetilde{\mathbf{U}}^{h(2)} \times_3 \widetilde{\mathbf{U}}^{h(3)} \tag{3.4}$$

Because each individual coefficient can be estimated separately, we have

$$\hat{y}^h_{r,s,t} = \underset{y_{r,s,t}}{\operatorname{argmax}} P(y_{r,s,t}|\mathcal{Y}^l), \tag{3.5}$$

From the assumption of low-correlation between the coefficients in $\mathcal{Y}^l$, we can simplify this probability in (3.5) as

$$P(y_{r,s,t}|\mathcal{Y}^l) \approx P(y_{r,s,t}|y^l_{1,1,1})..P(y_{r,s,t}|y^l_{Q_1,Q_2,Q_3}). \tag{3.6}$$

We can also rewrite $[y^l_{r,s,t}]$ into a vector form as

$$P(y_r|\mathcal{Y}^l) \approx \prod_{p=1}^{Q_1Q_2Q_3} P(y_r|y^l_p). \tag{3.7}$$

We use Gaussian to model the probability in (3.7) as:

$$P(y_r|y^l_p) \approx c\exp\{-\frac{(y_r - w_{r,p}y^l_p)^2}{2}\}, \tag{3.8}$$

where c is a constant. This Gaussian model evaluates the weighted distance between the projection coefficients. Equation (3.7) can be rewritten as

$$P(y_r|\mathcal{Y}^l) \approx c\exp\{-\sum_{p=1}^{Q_1Q_2Q_3} \frac{(y_r - w_{r,p}y^l_p)^2}{2}\}. \tag{3.9}$$

The Maximum Likelihood estimate of (3.9) is given by

$$\hat{y}^h_r = \underset{y_r}{\operatorname{argmax}} \log P(y_r|\mathcal{Y}^l). \tag{3.10}$$

We can express in a linear regression model as [76]

$$\hat{y}^h_r = \sum_{p=1}^{Q_1Q_2Q_3} w'_{r,p}y^l_p, \tag{3.11}$$

where

$$w'_{r,p} = \frac{w_{r,p}}{Q_1Q_2Q_3}. \tag{3.12}$$

We can calculate the value of $w'_{r,p}$ from the HR and LR of training sets. Each training image provides one equation to find $w'_{r,p}$ $(p = 1, ..., Q_1Q_2Q_3)$ and the column vector formed by

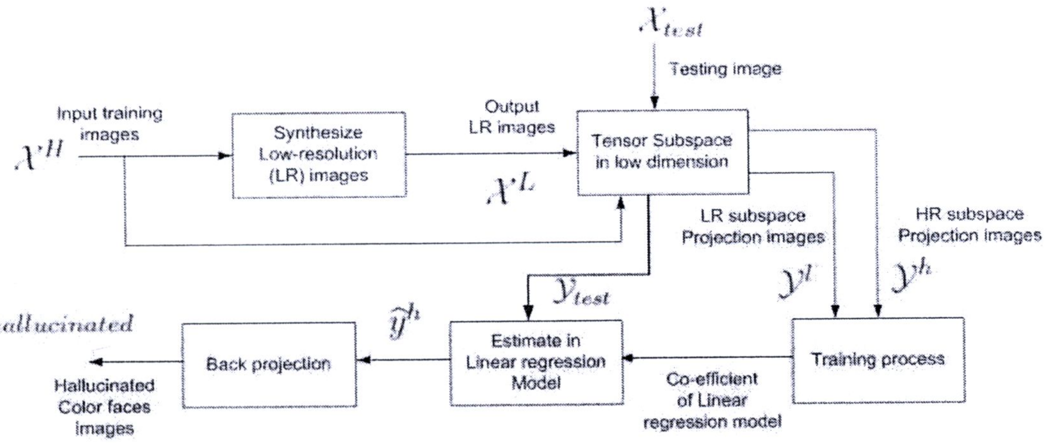

Figure 3.1: block diagram of color face hallucination with the linear regression model in MPCA

the rth projection coefficients for the HR images be $\mathbf{y}_r^h$. In the same way, the column vector formed by the pth projection coefficients for the corresponding LR set be $\mathbf{y}_p^l$ and we can state as:

$$\mathbf{y}_r^h = [\mathbf{y}_1^l, ..., \mathbf{y}_{Q_1Q_2Q_3}^l]\mathbf{w}_r, \tag{3.13}$$

where

$$\mathbf{w}_r = [w'_{r,1}, ..., w'_{r,Q_1Q_2Q_3}]^T. \tag{3.14}$$

After that $\mathbf{w}_r$ can be given by an Least-Square (LS) estimate:

$$\mathbf{w}_r = [\mathbf{y}_1^l, ..., \mathbf{y}_{Q_1Q_2Q_3}^l]^+\mathbf{y}_r^h. \tag{3.15}$$

3.2 Color Face Super-Resolution Based on Tensor Patch Method

In this section, we apply HOSVD tensor patch within the well-known framework for color face hallucination. A tensor structure provides a powerful mechanism to incorporate information and interaction of these image ensembles of multiple modalities at different resolutions. More precisely, given a training dataset of high-resolution face images, we blur and subsample them with different Gaussian filters and sub-sampling factors, while keeping the image size unchanged, so to generate a set of low-resolution training face images. To further improve the modeling accuracy, we uniformly decompose these face images into overlapped image blocks, and then obtain a hierarchical ensemble containing block-wise face images at low- and high-resolution. With these training data in place, we can construct a seventh-order tensor $\mathcal{D}$. We use HOSVD to decompose $\mathcal{D}$ into

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_{idens} \times_2 \mathbf{U}_{pixel1} \times_3 \mathbf{U}_{pixel2} \times_4 \mathbf{U}_{patch1} \times_5 \mathbf{U}_{patch2} \times_6 \mathbf{U}_{color} \times_7 \mathbf{U}_{resos} \tag{3.16}$$

where tensor $\mathcal{D}$ groups these block-wise training images into a tensor structure, and the core tensor $\mathcal{Z}$ governs the interactions between the seven mode factors. In (3.16), mode matrix $\mathbf{U}_{idens}$ spans the parameter space of identity, $\mathbf{U}_{pixel1}$ and $\mathbf{U}_{pixel2}$ span the space of pixel in x-axis and y-axis, $\mathbf{U}_{patch1}$ and $\mathbf{U}_{patch2}$ span the space of patch in x-axis and y-axis, $\mathbf{U}_{color}$ span the spaces of color space and $\mathbf{U}_{resos}$ span the space of resolution.

To model high-resolution details for the purpose of face hallucination, we uniformly decompose the low- and high-resolution face images into small overlapped patches, and perform tensor modeling at patch level. We can hallucinate high-resolution image data with all the decomposed patches. The final high-resolution face images are compositions of their corresponding overlapped small patches.

We suppose that $\mathbf{H}_1$ is the high-resolution color images, $\mathbf{S}_1$ is its low-resolution correspondences to be synthesized and $\mathbf{L}_1$ is any low-resolution color face input images. The task comes as finding the maximum *a posteriori* (MAP) estimation of $\mathbf{H}_1$ given $\mathbf{L}_1$ which can be formulated as

$$\{\mathbf{H}_{1MAP}\} = \text{argmax}_{\mathbf{H}_1,\mathbf{S}_1} \log P(\mathbf{H}_1, \mathbf{S}_1|\mathbf{L}_1) \tag{3.17}$$

By applying Bayes rule, we have

$$P(\mathbf{H}_1, \mathbf{S}_1|\mathbf{L}_1) = P(\mathbf{H}_1|\mathbf{S}_1, \mathbf{L}_1)P(\mathbf{S}_1|\mathbf{L}_1) \tag{3.18}$$

During the sequential processes of our face hallucination, the high-resolution face image is independently reconstructed. Based on the synthesized low-resolution image, the above expression can be state as

$$\begin{aligned} P(\mathbf{H}_1, \mathbf{S}_1|\mathbf{L}_1) &= P(\mathbf{H}_1|\mathbf{S}_1)P(\mathbf{S}_1|\mathbf{L}_1) \\ &= P(\mathbf{S}_1|\mathbf{H}_1)P(\mathbf{H}_1)P(\mathbf{S}_1|\mathbf{L}_1). \end{aligned} \tag{3.19}$$

The high-resolution image is naturally composed from the two part:

$$\mathbf{H} = \mathbf{H}^{lm} + \mathbf{H}^{h}, \tag{3.20}$$

where $\mathbf{H}^{lm}$ represents face images containing low- and middle-frequency information, and $\mathbf{H}^{h}$ contains high-frequency part. Since $\mathbf{H}^{lm}$ contributes the main part of after blurring and subsampling, then the probability $P(\mathbf{S}|\mathbf{H})$ can be approximated as $P(\mathbf{S}|\mathbf{H}^{lm})$. Based on (3.20), we also have $P(\mathbf{H}) = P(\mathbf{H}^{h}|\mathbf{H}^{lm})P(\mathbf{H}^{lm})$, and the estimation of $\mathbf{H}$ given $\mathbf{H}^{lm}$ is equivalent to the estimation of $\mathbf{H}^{h}$ given $\mathbf{H}^{lm}$, we then reformulate probability $P(\mathbf{S}_1|\mathbf{H}_1)P(\mathbf{H}_1)$ as

$$\begin{aligned} P(\mathbf{S}_1|\mathbf{H}_1)P(\mathbf{H}_1) &= P(\mathbf{S}_1|\mathbf{H}_1^{lm})P(\mathbf{H}_1^{h}|\mathbf{H}_1^{lm})P(\mathbf{H}_1^{lm}). \\ &= P(\mathbf{H}_1^{lm}|\mathbf{S}_1)P(\mathbf{H}_1|\mathbf{H}_1^{lm}). \end{aligned} \tag{3.21}$$

We can rewrite probability $P(\mathbf{S}_1|\mathbf{L}_1)$ as

$$P(\mathbf{S}_1|\mathbf{L}_1) = P(\mathbf{L}_1|\mathbf{S}_1)P(\mathbf{S}_1). \tag{3.22}$$

Probabilities $P(\mathbf{L}_1|\mathbf{S}_1)P(\mathbf{S}_1)$, $P(\mathbf{H}_1^{lm}|\mathbf{S}_1)$ and $P(\mathbf{H}_1|\mathbf{H}_1^{lm})$ sequentially constrain $\mathbf{S}_1$, $\mathbf{H}_1^{lm}$ and $\mathbf{H}_1$. This leads to a two-step sequential solution. In the first step, by using a global image-based tensor, we can synthesize the low-resolution $\mathbf{S}_1$. In the second step, after obtaining $\mathbf{S}_1$, the $\mathbf{H}_1^{lm}$ and $\mathbf{H}_1$ containing low-frequency, middle-frequency and high-frequency information can be computed using the local patch-based tensor. In addition, the final high-resolution $\mathbf{H}_1$ is computed by maximizing $P(\mathbf{H}_1|\mathbf{H}_1^{lm})$.

3.2.0.1 Global Low-Resolution Color Face Image Synthesis

In this section, the synthesis $\mathbf{S}_1$ is computed by maximizing probability $P(\mathbf{L}_1|\mathbf{S}_1)P(\mathbf{S}_1)$. Since $\mathbf{L}_1$ and $\mathbf{S}_1$ are the low-resolution given and synthesized face images with the same modality, we regard their relationship as Gaussian

$$P(\mathbf{L}_1|\mathbf{S}_1) = \frac{1}{f} exp\{-||\mathbf{S}_1 - \mathbf{L}_1||^2/\lambda\} \tag{3.23}$$

where f is a normalization constant and λ scales the variance.

In (3.16), if we index into its basis subtensor at a particular modality m^1, then the subtensor containing the individual image data as in (3.16) can be approximated by $\mathcal{G} = \mathcal{Z}_G \times_1 \mathbf{U}_{idens} \times_2 \mathbf{U}_{pixel1} \times_3 \mathbf{U}_{pixel2}$ and we get $\mathcal{G}_{m^1} = \mathcal{B}_{\mathcal{G}_{m^1}} \times_1 \mathbf{V}^T$. We unfold it into matrix representation and it becomes $\mathbf{G}_{m^1}^{(1)T} = \mathbf{B}_{\mathbf{G}_{m^1}^{T(1)}} \mathbf{V}$. Suppose $\mathbf{G}_{m^1}^{(1)T}$ correspond to color face images $\mathbf{S}_1$, then we substitute for $\mathbf{S}_1$ in (3.23) resulting in

$$P(\mathbf{L}_1|\mathbf{S}_1) = \frac{1}{f} exp\{-||\mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}} \mathbf{V} - \mathbf{L}_1||^2/\lambda\} \tag{3.24}$$

In reality the given low-resolution $\mathbf{L}_1$ and $\mathbf{S}_1$ synthesized have the same modality. By setting $\mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}} \mathbf{V} = \mathbf{L}_1$, we maximize (3.24) and approximately compute

$$\mathbf{V} = (\mathbf{B}_{\mathbf{G}_{m^1}^{(1)}} \mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}})^{-1} \mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}} \mathbf{L}_1 \tag{3.25}$$

where $(\mathbf{B}_{\mathbf{G}_{m^1}^{(1)}} \mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}})^{-1} \mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}}$ is the pseudoinverse of $\mathbf{B}_{\mathbf{G}_{m^1}^{(1)T}}$.

3.2.0.2 Local Patch-Based and High-Frequency Residue Recovery in Color Face Image Hallucination

To obtain their hallucinated high-resolution, we maximize $P(\mathbf{H}_1^{lm}|\mathbf{S}_1)$ using the local patch-based multiresolution tensor. The inference of $\mathbf{H}_1^{lm}$ from S_1 is independent. In the following, we take $\mathbf{H}_1^{lm}$ as an example to illustrate this second process.

Since the training local multiresolution tensor is constructed from small overlapped patches, we decompose the synthesized $\mathbf{S}_1$ uniformly in the same way as decomposing training data, and factorize the likelihood $P(\mathbf{H}_1^{lm}|\mathbf{S}_1)$ at patch level as

$$P(\mathbf{H}_1^{lm}|\mathbf{S}_1) = \prod_{p_1,p_2=1}^{N} P(\mathbf{H}_{1_{p_1,p_2}}^{lm}|\mathbf{S}_{1_{p_1,p_2}}). \tag{3.26}$$

Assuming $\mathbf{A}$ is the blurring and subsampling operator connecting $\mathbf{H}^{lm}_{1_{p_1,p_2}}$ and $\mathbf{S}_{1_{p_1,p_2}}$ in an imaging observation model, we regard these processes as Gaussian, therefore

$$P(\mathbf{H}^{lm}_1|\mathbf{S}_1) = \prod_{p_1,p_2=1}^{N} \frac{1}{w} exp\{-||\mathbf{A}\mathbf{H}^{lm}_{1_{p_1,p_2}} - \mathbf{S}_{1_{p_1,p_2}}||^2/\beta\} \tag{3.27}$$

where w is a normalization constant and β scales the variance.

Suppose the local multiresolution tensor in (3.16) has a basis tensor

$$\mathcal{B}_{\mathcal{L}} = \mathcal{Z}_{\mathcal{L}} \times_2 \mathbf{U}_{pixel1} \times_3 \mathbf{U}_{pixel2} \times_4 \mathbf{U}_{patch1} \times_5 \mathbf{U}_{patch2} \times_6 \mathbf{U}_{color} \times_7 \mathbf{U}_{resos}. \tag{3.28}$$

We index into this basis tensor at a particular resolution r and patch position p_1 and p_2, yielding a basis subtensor

$$\mathcal{B}_{\mathcal{L}_{r,p_1,p_2}} = \mathcal{Z}_{\mathcal{L}} \times_2 \mathbf{U}_{pixel1} \times_3 \mathbf{U}_{pixel2} \times_6 \mathbf{U}_{color} \times_4 \mathbf{V}^T_{p_1} \times_5 \mathbf{V}^T_{p_2} \times_7 \mathbf{V}^T_r. \tag{3.29}$$

Then as described in Section (2), the subtensor containing the pixel data for that particular patch can be approximated as $\mathcal{D}_{r,p_1,p_2} = \mathcal{B}_{\mathcal{L}_{r,p_1,p_2}} \times_1 \mathbf{V}^T$, and its unfolded representation is $\mathbf{D}^{(1)T}_{r,p_1,p_2} = \mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}} \mathbf{V}$. Similarly, we can obtain a subtensor for resolution r' of the same patch position, which is $\mathbf{L}^{(1)T}_{r',p_1,p_2} = \mathbf{B}_{\mathbf{L}^{(1)T}_{r',p_1,p_2}} \widetilde{\mathbf{V}}$. Suppose $\mathbf{L}^{(1)T}_{r,p_1,p_2}$ and $\mathbf{L}^{(1)T}_{r',p_1,p_2}$ correspond to $\mathbf{S}_{1_{p_1,p_2}}$ and $\mathbf{H}^{lm}_{1_{p_1,p_2}}$, respectively; we substitute them in (3.27) as

$$P(\mathbf{H}^{lm}_1|\mathbf{S}_1) = \prod_{p_1,p_2=1}^{N} \frac{1}{w} exp\{-||\mathbf{A}\mathbf{B}_{\mathbf{L}^{(1)T}_{r',p_1,p_2}} \widetilde{\mathbf{V}} - \mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}} \mathbf{V}||^2/\beta\}. \tag{3.30}$$

We optimize the parameter $\widetilde{\mathbf{V}}$ based on the construction properties of the local multiresolution patch tensor, which suggests that the relation between $\mathbf{B}_{\mathbf{L}^{(1)T}_{r',p_1,p_2}} \widetilde{\mathbf{V}}$ and $\mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}} \mathbf{V}$ observes a basic imaging observation model through the blurring and subsampling operator $\mathbf{A}$. This is consistent with the uniqueness of the identity parameter vector in a tensor space as well. By setting $\widetilde{\mathbf{V}} = \mathbf{V}$, we can approximately compute $\mathbf{H}^{lm}_{1_{p_1,p_2}}$ as

$$\mathbf{H}^{lm}_{1_{p_1,p_2}} = \mathbf{B}_{\mathbf{L}^{(1)T}_{r',p_1,p_2}} \Psi \mathbf{S}_{1_{p_1,p_2}} \tag{3.31}$$

where Ψ is the pseudoinverse of $\mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}}$ and is equal to $(\mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}} \mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}})^{-1}\mathbf{B}_{\mathbf{L}^{(1)T}_{r,p_1,p_2}}$. After reconstructing all the patches at different positions, the final hallucinated color face image $\mathbf{H}^{lm}_1$ is simply a composition of the corresponding hallucinated small patches.

We recover the highest frequency part by patch learning from the high-resolution training data. The inference of $\mathbf{H}_1$ from $\mathbf{H}^{lm}_1$ is independent. In the following, we take $\mathbf{H}_1$ as an example to illustrate how to hallucinate the final high-resolution face images.

We use a MRFs to model the $\mathbf{H}_1$ to be inferred. By decomposing into $\mathbf{H}^{lm}_1$ square patches

$$\begin{aligned} P(\mathbf{H}_1|\mathbf{H}^{lm}_1) &= P(\mathbf{H}^{lm}_1|\mathbf{H}_1)P(\mathbf{H}_1) \\ &= \prod_{q=1}^{Q} P(\mathbf{H}^{lm}_{1q}|\mathbf{H}_{1q})P(\mathbf{H}_1). \end{aligned} \tag{3.32}$$

The difference between $\mathbf{H}_1$ and $\mathbf{H}_1^{lm}$ is the high-frequency band information. Since the high-frequency information depends on the lower-frequency band, we use the Laplacian image $\mathbf{L}_{\mathbf{H}_1^{lm}}$ of $\mathbf{H}_1^{lm}$ to represent the middle-frequency band. To infer $\mathbf{H}_1$, we use the sum of squared differences of Laplacian images as metrics to model $\prod_{q=1}^{Q} P(\mathbf{H}_{1q}^{lm}|\mathbf{H}_{1q})$ as

$$\prod_{q=1}^{Q} P(\mathbf{H}_{1q}^{lm}|\mathbf{H}_{1q})\alpha \prod_{q=1}^{Q} exp-||\mathbf{L}_{\mathbf{H}_{1q}^{lm}} - \mathbf{L}_{\mathbf{H}_{1q}^{(t)}}||^2 \tag{3.33}$$

where $\mathbf{L}_{\mathbf{H}_{1q}^{(t)}}$ are the Laplacian images from high-resolution training face images. Comparing the Laplacian images $\mathbf{L}_{\mathbf{H}_{1q}^{lm}}$ with $\{\mathbf{L}_{\mathbf{H}_{1q}^{(t)}}\}_{t=1}^{T}$ from the training dataset, the patch $\mathbf{H}_{1q}^{(t)}$ with $\mathbf{L}_{\mathbf{H}_{1q}^{(t)}}$ closest to $\mathbf{L}_{\mathbf{H}_{1q}^{lm}}$ is the most probable to be chosen as $\mathbf{H}_{1q}$. Since we model the high-resolution image as a MRFs, based on the HammersleyClifford theorem, $P(\mathbf{H}_1)$ is a product $\prod_{\mathbf{H}_{1q},\mathbf{H}_{1\bar{q}}} \Phi(\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}})$ of compatibility functions $\Phi(\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}})$ over all neighboring pairs, where $\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}}$ are one of the neighboring patch pairs in a 4-neighbor system.

The compatibility function $\Phi(\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}})$ is defined using the similarity of pixel values on the overlapping area of the neighboring patches:

$$\Phi(\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}})\alpha exp\{-||O_{\mathbf{H}_{1q}} - O_{\mathbf{H}_{1\bar{q}}}||^2\}, \tag{3.34}$$

where $O_{\mathbf{H}_{1q}}$ denotes the pixels of patch $\mathbf{H}_{1q}$ overlapping with neighboring patch $\mathbf{H}_{1\bar{q}}$, and vice versa for $O_{\mathbf{H}_{1\bar{q}}}$. We illustrate this 4-neighbor system and the corresponding patch overlapping relations, then $\mathbf{H}_1$ estimated as

$$\text{argmax}_{\mathbf{H}_1} \prod_{q=1}^{Q} P(\mathbf{H}_{1q}^{lm}|\mathbf{H}_{1q}) \prod_{(q,\bar{q})} \Phi(\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}}). \tag{3.35}$$

Solving probabilistic (3.35) to obtain $\mathbf{H}_1$ is not a trivial task. We use the iterated conditional modes (ICM) algorithm [77]. More specifically, we maximize $P(\mathbf{H}_{1q}^{lm}|\mathbf{H}_{1q})$ for all patch positions $q \in \{1, ..., Q\}$ to yield the initial maximum likelihood estimate of . Based on this initial estimate, we then pick a random patch position q and update the estimate of $\mathbf{H}_{1q}$ using the current estimates of its neighbors $\mathbf{H}_{1\bar{q}}$ by maximizing $P(\mathbf{H}_{1q}^{lm}|\mathbf{H}_{1q}) \prod_{(q,\bar{q})} \Phi(\mathbf{H}_{1q}, \mathbf{H}_{1\bar{q}})$. We repeat this random patch selection and updating process until converging to the final high-resolution image $\mathbf{H}_1$.