# FACE HALLUCINATION VIA SPARSE CODING

*Jianchao Yang, Hao Tang, Yi Ma, Thomas Huang*

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
{jyang29,haotang2,yima,huang}@uiuc.edu

## ABSTRACT

In this paper, we address the problem of hallucinating a high resolution face given a low resolution input face. The problem is approached through sparse coding. To exploit the facial structure, Non-negative Matrix Factorization (NMF) [1] is first employed to learn a localized part-based subspace. This subspace is effective for super-resolving the incoming low resolution face under reconstruction constraints. To further enhance the detailed facial information, we propose a local patch method based on sparse representation with respect to coupled overcomplete patch dictionaries, which can be fast solved through linear programming. Experiments demonstrate that our approach can hallucinate high quality super-resolution faces.

***Index Terms—*** Super-resolution, face hallucination, non-negative matrix factorization, sparse representation, sparse coding.

## 1. INTRODUCTION

In most surveillance scenarios, there is a large distance between the camera and the objects of interest in the scene, usually resulting in low resolution of these objects. For tasks such as automatic face recognition and identification, it is often needed to enhance the resolution of the faces. Numerous super-resolution algorithms for generic images have been proposed in the literature [11], [10], [4], [8], [13], [5], which either achieve a Maximum *a Posteriori* (MAP) solution combing multiple frames under reconstruction constraints or generate a high resolution image from a single low resolution input using priors learned from local patch pairs. However, without consideration on special characteristics of face images, these algorithms are not so efficient when applied to very low resolution faces.

Baker and Kanade [3] started the pioneering work on face hallucination. As a heuristic method, the gradient pyramid-based prediction cannot model the face priors very well, and the pixels are predicted individually, causing discontinuity and artifacts. Liu et al. [12] proposed a two-step statistical approach integrating the global PCA model and a local patch model. Although the algorithm yields good results, it uses the holistic PCA model tending to render results like the mean face and the probabilistic local patch model is very complicated. Wei Liu et al. [2] proposed a new approach based on TensorPatch and residue compensation. Adding more details to the face, the algorithm also results in more artifacts.

In this paper, we propose a novel approach to the face hallucination problem through sparse coding. Non-negative Matrix Factorization is used to learn a localized parts-based representation, which is believed to be the principles of human learning by psychologists and physiologists. Then with coupled overcomplete dictionaries, the local patch based method from sparse representation is used to further enhance the resolution.

## 2. SPARSE CODING USING NON-NEGATIVE MATRIX FACTORIZATION

In face hallucination, the most frequently used subspace method for modeling the human face is PCA, which chooses a new coordinate system such that the variances of the dataset are preserved orderly. However, the PCA bases are holistic, making it unstable to occlusions. Compared to NMF, the reconstruction results of PCA are not that intuitive and hard to interpret as PCA allows subtractive combinations of the basis images.

Even though faces are objects with lots of variance, they are made up of several relatively independent parts such as eyes, eyebrows, noses, mouths, checks and chins. The idea behind Non-negative Matrix Factorization (NMF) [1] is to extract these relevant parts and find an additive combination of these local features, which is inspired by psychological and physiological principles assuming that humans learn objects in part-based owner. To find such a part-based subspace, NMF is formulated as the following optimization problem:

$$\begin{aligned} &\arg\min_{W,H} \|D - WH\|_2^2 \\ &s.t. \quad W \geq 0, H \geq 0, \end{aligned} \tag{1}$$

where $D \in \Re^{n\times m}$ is the data matrix, $W \in \Re^{n\times r}$ is the basis matrix and $H \in \Re^{r\times m}$ is the coefficient matrix. The number

This work was supported in part by the U.S. Government VACE Program.

of the bases $r$ can be chosen as $n*m/(n+m)$ which is smaller than $n$ and $m$, meaning a more compact representation. It can be shown that the solution to (1) can be achieved through the following updata rules:

$$
\begin{aligned}
H_{ij} &\longleftarrow H_{ij}\frac{(W^T D)_{ij}}{(W^T W H)_{ij}} \\
W_{ti} &\longleftarrow W_{ti}\frac{(DH^T)_{ti}}{(WHH^T)_{ki}},
\end{aligned}
\tag{2}
$$

where $1 \leq i \leq r$, $1 \leq j \leq m$ and $1 \leq t \leq n$.

Let $I_h$ and $I_l$ denote the high resolution and low resolution faces respectively. $I_l$ is obtained from $I_h$ by smoothing and downsampling. Write $I_h$ and $I_l$ as long vectors by stacking their columns, the image degradation process from high to low resolution can be formulated as $I_l = MI_h$, where $M$ is a matrix performing both blurring and downsampling. Given $I_l$, we can achieve the optimal solution for $I_h$ based on the Maximum *a Posteriori* (MAP) criterion,

$$
I_h^* = \arg\max_{I_h} p(I_l|I_h)p(I_h). \tag{3}
$$

Using the rules in (2), we can obtain the basis matrix $W$, which is composed of sparse bases. Let $\Omega$ denote the face subspace spanned by $W$. Then in the subspace $\Omega$, the super-resolution problem in (3) can be reformulated as:

$$
\begin{aligned}
& c^* = \arg\min_c \|MWc - I_l\|_2^2 + \lambda\rho(Wc) \\
& s.t. \quad c \geq 0,
\end{aligned}
\tag{4}
$$

where $\rho(Wc)$ is a prior term regularizing the high resolution solution, $c \in \Re^{r\times 1}$ is the coefficient vector in the subspace $\Omega$ for estimated the high resolution face, and $\lambda$ is a parameter used to balance the reconstruction fidelity and the penalty of the piror term. A lot of image priors have been proposed over the years to regularize those ill-posed inverse problems in image processing including super-resolution.

In this paper, we simply use a generic image prior requiring that the solution should be smooth. Let $\Gamma$ denote a matrix performing high-pass filtering. The final formulation for (4) is:

$$
\begin{aligned}
& c^* = \arg\min_c \|MWc - I_l\|_2^2 + \lambda\|\Gamma Wc\|_2 \\
& s.t. \quad c \geq 0.
\end{aligned}
\tag{5}
$$

The high resolution image $I_h$ is approximated by $Wc^*$. However, the prior term in (5) suppresses the high frequency components, resulting in over-smoothness in the solution image. We will rectify this in Sec. 3 using a local patch method based on sparse representation.

## 3. SPARSE CODING FROM SPARSE REPRESENTATION

In recent years, there has been a growing interest in the study of sparse representation of signals. Using an overcomplete dictionary that contains prototype signal-atoms, signals are represented as sparse linear combinations of these atoms. Specifically, let $\mathbf{D} \in \Re^{n\times K}$ be an overcomplete dictionary containing $K$ prototype signal-atoms, and suppose a signal $\mathbf{x} \in \Re^n$ can be represented as a sparse linear combination of these atoms. That is, $\mathbf{x} = \mathbf{D}\alpha_0$ where $\alpha_0 \in \Re^K$ is a vector with very few ($\ll K$) nonzero entries. In practice, we might observe only a small set of or corrupted measurements $\mathbf{y}$ of $\mathbf{x}$:

$$
\mathbf{y} \doteq L\mathbf{x} = L\mathbf{D}\alpha_0, \tag{6}
$$

where $L \in \Re^{k\times n}$ with $k \leq n$. In our super-resolution context, $\mathbf{x}$ is a high-resolution image patch, while $\mathbf{y}$ is its low-resolution counterpart (or features extracted from it). If the dictionary $\mathbf{D}$ is overcomplete, the equation $\mathbf{x} = \mathbf{D}\alpha_0$ is underdetermined for the unknown coefficients $\alpha_0$. The equation $\mathbf{y} = L\mathbf{D}\alpha_0$ is even more underdetermined. Nevertheless, under mild conditions, the sparsest solution $\alpha_0$ to this equation will be unique. Furthermore, if $\mathbf{D}$ satisfies an appropriate near-isometry condition, then for a wide variety of matrices $L$, any sufficiently sparse linear representation of a high-resolution image $\mathbf{x}$ in terms of $\mathbf{D}$ can be recovered from the low-resolution image.

Since the results given by (5) are smooth because of the prior term, we further use a local patch method to add more details. As in other patch-based methods, we divide the incoming low resolution image into overlapped patches and try to infer the high-resolution patch for each low-resolution patch from the input. Suppose we have low-resolution face images $\{I_l^{(i)}\}_{i=1}^{ns}$ and their high-resolution counterparts $\{I_h^{(i)}\}_{i=1}^{ns}$ as our training set, where $ns$ is the sample number. For each image pair in the training set, randomly sample patch pairs in the same locations from them. We arrange all these patches into two matrices: $\mathbf{D}_\ell = [y^{(1)}, ..., y^{(N)}]$ and $\mathbf{D}_\hbar = [x^{(1)}, ..., x^{(N)}]$, where $y^{(i)}$ is the vector representation of the $i$-th low resolution patch, and $x^{(i)}$ is the vector representation of the corresponding high resolution patch. $\mathbf{D}_\ell$ and $\mathbf{D}_\hbar$ form the two dictionaries we use for our sparse representation algorithm. We subtract the mean pixel value for each patch, so that the dictionaries represent image patterns more succinctly.

For each input low-resolution patch $y$, we find a sparse representation for it with respect to the low-resolution dictionary $\mathbf{D}_\ell$. The corresponding high-resolution patches $\mathbf{D}_\hbar$ will be combined according to these sparse coefficients to generate the output high-resolution patch $x$. The problem of finding the sparsest representation of $y$ can be formulated as:

$$
\min \|\alpha\|_0 \quad \text{s.t.} \quad \|F\mathbf{D}_\ell\alpha - F\mathbf{y}\|_2^2 \leq \epsilon, \tag{7}
$$

where $\|\cdot\|_0$ denotes the zero norm (the number of nonzero entries of the vector), and $F$ is a feature extraction operator. In this paper, $F$ is chosen as a gradient filter. The main role of $F$ in (7) is to provide a perceptually meaningful constraint on how closely the coefficients $\alpha$ must approximate $y$.

Although the optimization problem (7) is NP-hard in general, recent results [6] indicate that as long as the desired coefficients $\alpha$ are sufficiently sparse, they can be efficiently recovered by instead minimizing the $\ell^1$-norm, as follows:

$$\min \|\alpha\|_1 \quad \text{s.t.} \quad \|F\mathbf{D}_\ell\alpha - F\mathbf{y}\|_2^2 \le \epsilon. \tag{8}$$

Lagrange multipliers offer an equivalent formulation:

$$\min \ \eta\|\alpha\|_1 + \tfrac{1}{2}\|F\mathbf{D}_\ell\alpha - F\mathbf{y}\|_2^2, \tag{9}$$

where the parameter $\eta$ balances the sparsity of the solution and fidelity of the approximation to $y$. Notice that this is essentially a linear regression regularized with $\ell^1$-norm on the coefficients, known in the statistical literature as the Lasso [14].

Solving (9) individually for each patch however does not guarantee compatibility between adjacent patches. We enforce compatibility between adjacent patches using a one-pass algorithm similar to that of [9]. The patches are processed in raster-scan order in the image, from left to right and top to bottom. We modify (8) so that the super-resolution reconstruction $\mathbf{D}_\hbar\alpha$ of the patch $y$ is constrained to closely agree with the previously computed adjacent high-resolution patches. The resulting optimization problem is

$$\begin{aligned} \min \|\alpha\|_1 \quad \text{s.t.} \quad & \|F\mathbf{D}_\ell\alpha - F\mathbf{y}\|_2^2 \le \epsilon_1 \\ & \|P\mathbf{D}_\hbar\alpha - \mathbf{w}\|_2^2 \le \epsilon_2, \end{aligned} \tag{10}$$

where the matrix $P$ extracts the overlapped region between the current target patch and previously reconstructed high-resolution image, and $\mathbf{w}$ contains the values of the previously reconstructed high-resolution image on the overlap. The constrained optimization (10) can be similarly reformulated as:

$$\min \ \eta\|\alpha\|_1 + \tfrac{1}{2}\|\tilde{\mathbf{D}}\alpha - \tilde{\mathbf{y}}\|_2^2, \tag{11}$$

where $\tilde{\mathbf{D}} = \begin{bmatrix} F\mathbf{D}_\ell \\ \beta P\mathbf{D}_\hbar \end{bmatrix}$ and $\tilde{\mathbf{y}} = \begin{bmatrix} F\mathbf{y} \\ \beta\mathbf{w} \end{bmatrix}$. The parameter $\beta$ controls the tradeoff between matching the low-resolution input and finding a high-resolution patch that is compatible with its neighbors. In our experiments, we simply set $\beta = 1$. Given the optimal solution $\alpha^*$ to (11), the high-resolution patch can be reconstructed as $\mathbf{x} = \mathbf{D}_\hbar\alpha^*$.

## 4. FACE HALLUCINATION FROM SPARSE CODING ALGORITHM

In this section we summarize our face hallucination algorithm. In the first step, using the sparse subspace $\Omega$, we can recover the global face structure and main local features of the target high resolution image. However, the result image is over smoothed from the first step because of the smoothness prior term. To further enhance the local detail information, we employ the sparse representation technique with respect to coupled dictionaries $\mathbf{D}_\ell$ and $\mathbf{D}_\hbar$ for each input patch. The complete framework of our algorithm is summarized as Algorithm 1.

**Algorithm 1** (Face Hallucination via Sparse Coding).

1: Input: sparse basis matrix $W$, training dictionaries $\mathbf{D}_\hbar$ and $\mathbf{D}_\ell$, a low-resolution image $I_l$.

2: Find a smooth high resolution face $Y$ from the subspace spanned by $W$ through:

- Solve the optimization problem in (5): $\arg\min_c \|MWc - I_l\|_2 + \lambda\|\Gamma Wc\|_2 \quad s.t. \quad c \ge 0$.
- $Y = Wc^*$.

3: For each patch $y$ of $Y$, taken starting from the upper-left corner with 1 pixel overlap in each direction,

- Solve the optimization problem with $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{y}}$ defined in (11): $\min \eta\|\alpha\|_1 + \frac{1}{2}\|\tilde{\mathbf{D}}\alpha - \tilde{\mathbf{y}}\|_2^2$.
- Generate the high-resolution patch $x = \mathbf{D}_\hbar\alpha^*$. Put the patch $x$ into a high-resolution image $X^*$.

4: Output: super-resolution face $X^*$.

## 5. EXPERIMENT RESULTS

Our experiments were conducted on the face database FRGC Ver 1.0 [15]. All these face images were aligned by an automatic alignment algorithm using the eye positions, and then cropped to the size of $100 \times 100$ pixels. To obtain the sparse subspace $\Omega$ spanned by $W$, we selected 540 face images, covering both genders, different races and facial expressions (Figure 1). To prepare the coupled dictionaries needed by our sparse representation algorithm, we sampled approximately 100,000 patch pairs from the training images. The patches are of size $5 \times 5$ pixels. 30 outside faces were chosen as our testing cases, which were blurred and downsampled to the size of 25-by-25 pixels.

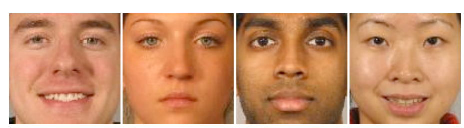

**Fig. 1**. Example training faces in our algorithm.

In our algorithm, there is only one parameter $\lambda$ (5) that we need to determine. Experimentally, we find that setting $\lambda = 0.005$ generally offers satisfactory results. We compare our algorithm with bicubic interpolation and backprojection [11]. The results are shown in Figure 2, which indicate that our method can generate much higher resolution faces. From columns 4 and 5, we can also see that the local patch method based on sparse respresentation further enhances the edges and textures.

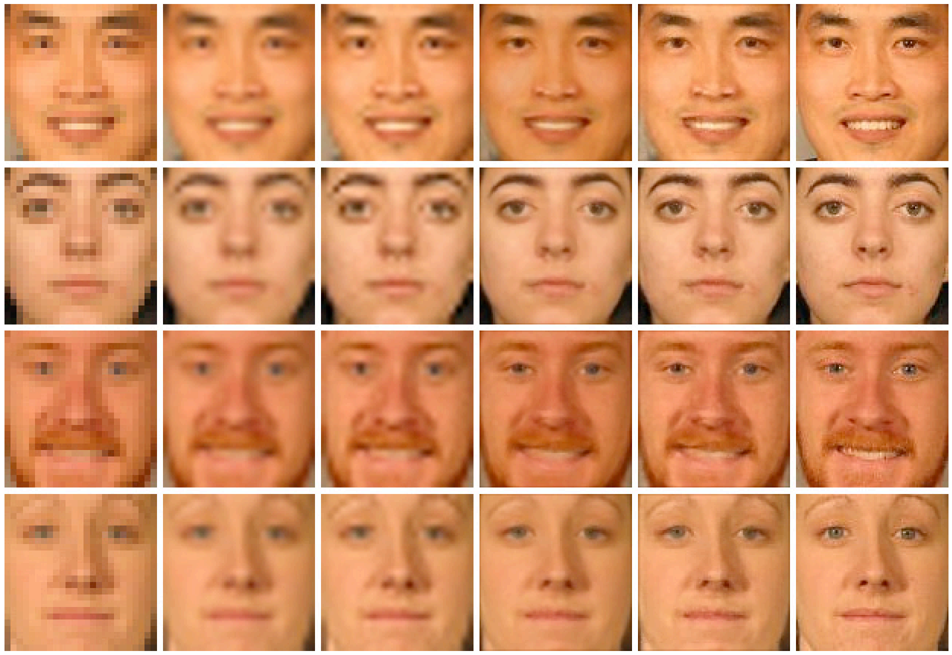

**Fig. 2**. Results of our algorithm compared to other methods. From left to right columns: low resolution input; bicubic interpolation; back projection; sparse coding via NMF followed by bilater filtering; sparse coding via NMF and Sparse Representation; Original.

## 6. CONCLUSION

In this paper, we propose a new method for face hallucination via sparse coding (sparse basis coding and sparse representation). Although we only use a small database and a simply face alignment algorithm, the results already reveal the potential of our algorithm for hallucinating faces. A larger training database and more complicated face alignments as in [12] and [2] will promise better results, and we leave that to our future work.

## 7. REFERENCES


[1] Lee, D.D., And Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature 401*, 6755 (October 1999), 788-791.

[2] Wei Liu, Dahua Lin, and Xiaoou Tang. Face Hallucinating Faces: TensorPatch Super-Resolution and Coupled Residue Compensation. *Proc. CVPR*,2006.

[3] S. Baker and T. Kanade. Hallucinating Faces. *IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.*

[4] D. Capel. Image mosaicing and super-resolution. Ph.D. Thesis, Department of Eng. Science, University of Oxford, 2001.

[5] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. *CVPR*, 2004.

[6] D. L. Donoho. For most large underdetermined systems of linear equations, the minimal $\ell^1$-norm solution is also the sparsest solution. Comm. on Pure and Applied Math, Vol. 59, No. 6, 2006.

[7] D. L. Donoho. For most large underdetermined systems of linear equations, the minimal $\ell^1$-norm near-solution approximates the sparsest near-solution. Preprint, accessed at http://www-stat.stanford.edu/ donoho/. 2004.

[8] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *IJCV*, 2000.

[9] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, Vol. 22, Issue 2, 2002.

[10] R.C. Hardie, K.J. Barnard, and E.A. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE TIP*, 1997.

[11] M. Irani and S. Peleg. Motion analysis for image enhancement: resolution, occulusion and transparency. *JVCI*, 1993.

[12] C. Liu, H. Y. Shum, and W. T. Freeman. Face hallucination: theory and practice. *IJCV*, Vol. 75, No. 1, pp. 115-134, October, 2007.

[13] J. Sun, N.-N. Zheng, H. Tao, and H. Shum. Image hallucination with primal sketch priors. *Proc. CVPR*, 2003.

[14] R. Tibshirani. Regression shrinkge and selection via the Lasso. J. Royal Statist. Soc B., Vol. 58, No. 1, pages 267-288, 1996.

[15] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek. Overview of Face Recognition Grand Challenge. *Proc. CVPR*, 2005.