MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

# A Two-dimensional Direct Combined Model for Facial Hallucination

Ching-Ting Tu[1] Kung-Hung Lin[1] Tyng-Luh Liu[2]
[1]Department of Computer Science and Information Engineering Tamkang University, Taiwan
[2]Institute of Information Science, Academia Sinica, Taiwan
cttu@mail.tku.edu.tw drhiromu@gmail.com liutyng@iis.sinica.edu.tw

## Abstract

*This study developed a face hallucination system based on a novel two-dimensional direct combined model (2DDCM) approach that employs a large collection of low-resolution/high-resolution facial pairwise training examples. The proposed 2DDCM approach achieves face hallucination by addressing three key issues. First, we directly combine each low-resolution and high-resolution pairwise image in a concatenated form in order to completely preserve their relationship. Second, images are formed as two–dimensional matrices instead of vectors in order to preserve the facial geometry. Third, both the vertical and the horizontal facial-geometry features are considered in 2DDCM approach. Experiments demonstrate our approach can synthesize high quality reconstructed facial images from given low-resolution images.*

## 1. Introduction

The need for face hallucination, i.e., reconstructing the high-resolution facial image from a given low-resolution image, is present in many computer vision and multimedia applications. For example, the task of recognizing faces in a video [1, 3] using reconstructed high resolution facial images may improve the facial recognition rate or at least aid the recognition task. Unlike other super resolution approaches proposed in [4, 8, 13], where the targeted images are without any particular structure, the images processed in the face hallucination framework consist of common facial structures (e.g. the eyes and nose). Consequently, we propose a two-dimensional direct combined model (2DDCM) approach for face hallucination with the help of a set of low-resolution/high-resolution facial pairwise training examples in order to learn facial structures.

Naturally, the effectiveness of face hallucination depends on the accuracy of the underlying transformation between the low- and the high-resolution facial images. Accordingly, the proposed 2DDCM approach addresses the following three distinguishing characteristics for the derived transformation of face hallucination: (1) Compared to existing independent model approaches [6, 9, 10], the relationship modeled by the combined formulation method [2, 5, 12] is more flexible and completely reveals their primary facial properties. Consequently, the 2DDCM approach directly couples each pairwise example in a combined formula in order to completely preserve their correlation. (2) Inspired by approaches in [11, 14], we apply the 2D matrix image representation in the proposed 2DDCM approach. Compared to the conventional pixel-based vector image representation, the 2D matrix image representation doesn't destroy facial structure and provides significant help for further analyzing the vertical and horizontal facial structures. (3) The pairwise training examples in 2D combination representations are then used for deriving the transformation of face hallucination. Such transformation synthesizes quality reconstructed faces by emphasizing the facial and the horizontal characteristics; these characteristics are important, but have been given less emphasis in previous literature, e.g. [5, 6, 7, 9, 10, 11].

## 2. Two-dimensional direct combined model

Our method for deriving the transformation of face hallucination starts with a novel representation, 2D combination formulation, of two related classes. In the current scenario, one class is the set of low-resolution facial images $L$, and the other class is the set of their corresponding high-resolution facial images, $H$. Assume there are $K$ training pairwise examples, $\{(l_1,h_1),(l_2,h_2),\ldots,(l_K,h_K)\}$, where each image of the pair is a random vector of the class $L$ or $H$. Consider now each image ($l_i$ or $h_i$) are $M \times N$ pixels ($l_i$ is up-sampled to the same size as the targeted high-resolution image $h_i$). To better preserve the geometric properties of faces, such as the symmetry of the facial structure or the relative sizes of facial features, we then define the $M \times N$ facial image as a $M \times N$ random matrix, i.e., $l_i$ is represented as a matrix form $A_i^L$ and the matrix form of $h_i$ is $A_i^H$. Further, to carefully preserve the pairwise correlation, the *i-th* pairwise training example, $(l_i, h_i)$, is formulated by 2D combination representation $A_i$ .with size $2M \times 2N$:

$$A_i = \begin{bmatrix} A_i^L & A_i^H \\ A_i^H & A_i^H \end{bmatrix}_{2MX2N}, \tag{1}$$

where upper $M$ rows of $A_i$ correspond to the low-resolution class, $L$, and the lower $M$ rows of $A_i$ correspond to the high-resolution class, $H$. The last $N$ columns of $A_i$ are the augmented matrix for the pair of ($A_i^H$, $A_i^H$), i.e. we constrict the image resolution of class $L$ to its high-resolution version in (1).

To better expose vertical and the horizontal characteristic features of these $K$ training examples, we extracted two feature spaces $U$ and $V$, where the variance of all training 2D combination matrices, $\{A_1, A_2, \ldots, A_K\}$ on these two spaces is maximized. Such covariance matrix is defined as:

$$C_{U,V} = \frac{1}{K}\sum_{i=1}^{K}[U^T \Delta A_i V][U^T \Delta A_i V]^T \tag{2}$$

where $\Delta A_i$ denotes the unbiased matrix of $A_i$, i.e. $\Delta A_i = A_i - \overline{A}$; $\overline{A}$ is the mean matrix of all training matrices.

Using the matric trace $tr(C_{U,V}) = tr(V^T \Delta A^T U U^T \Delta A V)$ $= tr(U^T \Delta A V V^T \Delta A^T U)$, spaces $U$ and $V$ in (2) are solved

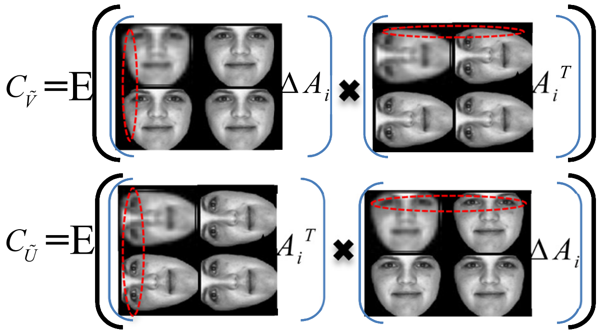

Figure 1. Illustrations of $C_{\tilde{V}}$ and $C_{\tilde{U}}$ in (5) and (6) with either $VV^T = I$ (in (5)) or $UU^T = I$ (in (6)). ($E$ denotes the empirical expectation).

by the following flip-flop procedure:

**U-step:** Let $\tilde{V}$ be the current estimated $V$-space, rewrite the objective function, and $U$ is found using the following optimization formula:

$$\max_{UU^T=I} \sum_{i=1}^{K} tr(U^T \Delta A_i \tilde{V}\tilde{V}^T \Delta A_i^T U) \tag{3}$$

**V-step:** Similarly, with the estimated $\tilde{U}$ -space, we can solve for $V$-space by:

$$\max_{VV^T=I} \sum_{i=1}^{K} tr(V^T \Delta A_i \tilde{U}\tilde{U}^T \Delta A_i^T V) \tag{4}$$

The matrix trace is a linear function, thus the summation can be moved inside the trace in (3) and (4). Let $C_{\tilde{V}}$ denote the covariance matrix along the $\tilde{V}$ space:

$$C_{\tilde{V}} = \begin{pmatrix} C_{\tilde{V}}^{LL} & C_{\tilde{V}}^{LH} \\ C_{\tilde{V}}^{HL} & C_{\tilde{V}}^{HH} \end{pmatrix} = \frac{1}{K}\sum_{i=1}^{K} \Delta A_i \tilde{V}\tilde{V}^T \Delta A_i^T, \tag{5}$$

and analogously,

$$C_{\tilde{U}} = \begin{pmatrix} C_{\tilde{U}}^{LL} & C_{\tilde{U}}^{LH} \\ C_{\tilde{U}}^{HL} & C_{\tilde{U}}^{HH} \end{pmatrix} = \frac{1}{K}\sum_{i=1}^{K} \Delta A_i^T \tilde{U}\tilde{U}^T \Delta A_i, \tag{6}$$

, where $C^{LL}$ $C^{HH}$ are the covariance matrix of class $L$ and class $H$ respectively and $C^{LH} = (C^{LH})^T$ is the cross-covariance matrix between these two classes.

We accordingly solve the optimal projection axes, $U$ space, in (3) by applying the singular value decomposition (SVD) process on $C_{\tilde{V}}$, where $\{U_1,\ldots,U_d\}$ are the orthonormal eigenvectors corresponding to the leading $d$ largest eigenvalues $\{\lambda_U^1,\ldots,\lambda_U^d\}$. Similarly, $\{V_1,\ldots,V_{d'}\}$ in (4) are the orthonormal eigenvectors of $C_{\tilde{U}}$ corresponding to the leading $d'$ largest eigenvalues $\{\lambda_V^1,\ldots,\lambda_V^{d'}\}$. That is, $V$ and $U$ spaces are solved by the following two eigenvalue formula:

$$\begin{aligned} \begin{pmatrix} C_{\tilde{V}}^{LL} & C_{\tilde{V}}^{LH} \\ C_{\tilde{V}}^{HL} & C_{\tilde{V}}^{HH} \end{pmatrix} &\approx \begin{bmatrix} U_L \Sigma_d^2 U_L^T & U_L \Sigma_d^2 U_H^T \\ U_H \Sigma_d^2 U_L^T & U_H \Sigma_d^2 U_H^T \end{bmatrix} \\ \begin{pmatrix} C_{\tilde{U}}^{LL} & C_{\tilde{U}}^{LH} \\ C_{\tilde{U}}^{HL} & C_{\tilde{U}}^{HH} \end{pmatrix} &\approx \begin{bmatrix} V_L \Sigma_d'^2 V_L^T & V_L \Sigma_d'^2 V_H^T \\ V_H \Sigma_d'^2 V_L^T & V_H \Sigma_d'^2 V_H^T \end{bmatrix} \end{aligned} \tag{7}$$

where we denote $U = [U_L^T \ U_H^T]^T$ and $V = [V_L^T \ V_H^T]^T$. The two-dimensional direct combined model is $U$ and $V$.

The roles the two-dimensional direct combined models, $V$ and $U$, are shown in Figure 1:

- $C_{\tilde{V}}$ in (5) is the within-individual covariance matrix between column vectors. That is, the extracted $U$ vectors give the major components of the facial vertical structures.
- Similarly, $C_{\tilde{U}}$ is the within-individual covariance matrix between row vectors. It indicates feature extraction is performed to emphasize facial horizontal structures shared by the columns of the training pairs.

In summary, incorporating by the 2DDCM, $V$ and $U$, the structurally-meaningful facial characteristics of the training pairs are extracted.

## 3. 2DDCM-based transformation

Once the 2DDCM are established, we then learn the transformation from class $L$ to class $H$. As with typical eigen-based approaches, any pairwise facial example, $A$, can be reconstructed by:

$$\begin{aligned} P(A \mid w_U) &\propto \exp\{-\left\|A - (Uw_U + \bar{A})\right\|^2 / \sigma_U^2\} \\ P(A \mid w_V) &\propto \exp\{-\left\|A^T - (Vw_V + \bar{A}^T)\right\|^2 / \sigma_V^2\} \end{aligned} \tag{8}$$

where $\sigma_V^2$ and $\sigma_U^2$ are variance of error term. The distribution of the weight vector $w_V$ or $w_U$ is also Gaussian, i.e.

$$P(w_U) \propto \exp\left\{-w_U^T \Lambda_U^{-1} w_U\right\};\ P(w_V) \propto \exp\left\{-w_V^T \Lambda_V^{-1} w_V\right\} \tag{9}$$

where $\Lambda_U = dig\{\lambda_U^1,\ldots,\lambda_U^d\}$ and $\Lambda_V = dig\{\lambda_V^1,\ldots,\lambda_V^{d'}\}$ are diagonal matrices; $\{\lambda_U^1,\ldots,\lambda_U^d\}$ are the eigenvalues of $C_{\tilde{V}}$, and $\{V_1,\ldots,V_{d'}\}$ are the eigenvalues of $C_{\tilde{U}}$.

From (8) and (9), we have the minimum mean-square-error (MMSE) estimator of $A^H \in H$:

$$\hat{A}^H(A^L) = \arg\min_{A^H} \iint (A^H - \hat{A}^H)^2 P(A,w)dAdW = \arg\min_{A^H} \int P(A^L \mid w)P(w) \int (A^H - \hat{A}^H)^2 P(\hat{A}^H \mid A^L, w)dAdW \tag{10}$$

,where $\hat{A}^H$ is the estimated vector of $A^H$; $P(A,w)$ is the joint probability of the observable matrix $A^L$, the unobservable matrix $A^H$, and the unobservable matrix $w$; $\hat{A}^H(A^L)$ is the estimated matrix $A^H$ for a given $A^L$; the result of the first equation is approximated by maximizing the second equation since the weight matrix $w$ is constricted by $A$.

We solve (10) by a two-step inference: First, estimate $w$ by maximizing $P(A^L \mid w)P(w)$:

$$\begin{cases} w_U = \arg\min_{w_U} \left\|\Delta A^L - U_L w_U + \lambda_U w_U{}^T \Lambda_U w_U\right\|^2 \\ w_V = \arg\min_{w_V} \left\|\Delta A^L - V_L w_V + \lambda_V w_V{}^T \Lambda_V w_V\right\|^2 \end{cases} \tag{11}$$

where $\Delta A^L$ is the unbiased matrix of $A^L$, and $\lambda$ is the predefined weights for (8) and (9). Since the objective function is a quadratic form, we get closed-form solutions:

$$\begin{cases} \hat{w}_U = \left(U_L{}^T U_L + \lambda_U \Lambda_U{}^{-1}\right)^{-1} U_L{}^T \Delta A^L \\ \hat{w}_V = \left(V_L{}^T V_L + \lambda_V V_U{}^{-1}\right)^{-1} V_L{}^T (\Delta A^L)^T \end{cases} \tag{12}$$

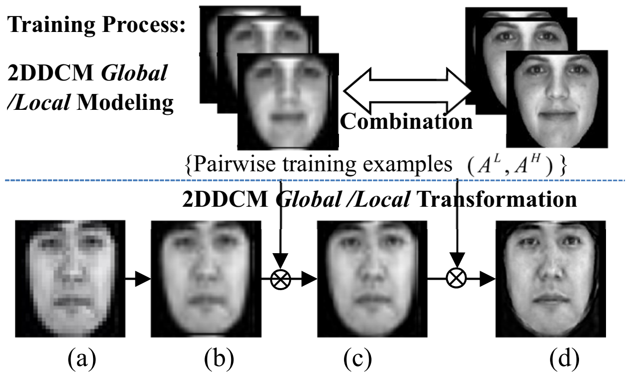


Figure 2. Framework of proposed face hallucination using 2DDCM. (a) low-resolution input. (b) result of (a) by Cubic B-Spline. (c) result by the proposed method after global stage (d) final output.

Second, the estimated $\hat{A}^H$ is considered as the expected matrix of the posterior probability of $A^H$ for a given observation $A^L$ and an estimated $\hat{w}$, i.e. $E[A^H \mid A^L, \hat{w}]$. Under the assumption that the joint distribution of $A^L$ and $A^H$ is a single Gaussian distribution, the MMSE estimator, e.g. (10), becomes:

$$\Delta\hat{A}^H(A^L) = C^{HL}(C^{LL})^{-1}\Delta A^L(\hat{w}) \qquad (13)$$

where $\Delta\hat{A}^H(A^L)$ is the unbiased estimated matrix of $A^H$ for a given $A^L$, $\Delta A^L(w)$ is the unbiased matrix of $A^L$ given $w$, the $C^{HL}$ is the cross-covariance matrix of class $H$ and class $L$, and $(C^{LL})^{-1}$ represents the inverse covariance matrix of $L$.

Finally, using the 2DDCM model, $U$ and $V$, in (7), the optimal $A^H$ by the 2DDCM transformation in (13) becomes:

$$\begin{cases} \Delta A_U^H = U_H\left(U_L{}^T U_L + \lambda_U \Lambda_U{}^{-1}\right)^{-1} U_L{}^T \Delta A^L \\ (\Delta A_V^H)^T = V_H\left(V_L{}^T V_L + \lambda_V \Lambda_V{}^{-1}\right)^{-1} V_L{}^T (\Delta A^L)^T \end{cases} . \qquad (14)$$

In summary, although $A^L$ and $A^H$ are initially combined in two-dimensional direct combined model, the transformation of 2DDCM algorithm seeks to estimate (separate) $A^H$ from a given $A^L$, as shown in (14). Such transformation is directly derived from the 2D learned feature spaces, $U$ and $V$, which facilitate exploring the meaningful relationship between the structures of classes $L$ and $H$.

## 4. Face hallucination using 2DDCM

This section introduces the framework based on the 2DDCM approach for face hallucination. As illustrated in Figure. 2, our face hallucination framework involves the global and local stages for reconstructing high-resolution facial images. The training procedure is described in Table 1, where the global stage consists of one block, $b_{num} = 1$, and the local stage of $b_{num} = 4$.

Table 1. Training procedure of 2DDCM alogrithm

**2DDCM Modeling ($b_{num}$)**:
1. Equally divide each training image $l_i$ or $h_i$ into $b_{num} \times b_{num}$ non-overlapping regions.
2. For each region $r = 1,..,b_{num} \times b_{num}$,
- **Initialize** $\tilde{U}$ and $\tilde{V}$ as identity matrices
- **Formulate** each patch-pair by 2D combination representation in (1)
- **Repeat** the following steps, at $t^{th}$ step
  (a) Compute $C_{\tilde{V}}^{r,t}$ by (5), update $\tilde{U}^{r,t} \leftarrow U$ by (7).
  (b) Compute $C_{\tilde{U}}^{r,t}$ by (6), update $\tilde{V}^{r,t} \leftarrow V$ by (7).
  (c) Compute the objective function $C_{U,V}^{r,t}$ in (2)
- **Stop when** $C_{U,V}^{r,t} - C_{U,V}^{r,t-1} < \varepsilon I$

**Return:** $b_{num} \times b_{num}$ $U$ and $V$ models.

In the reconstruction process, the low-resolution input image $l$ is first up-sampled to the target size and presented as an $M \times N$ image. The transformation of the 2DDCM for the testing process of face hallucination is described as shown in Table 2.

Table 2. Testing procedure of 2DDCM alogrithm

**2DDCM Transformation** ($b_{num}$):
1. Equally divide the up-sampled input $M \times N$ image, $A^L$, into $b_{num} \times b_{num}$ non-overlapping regions.
2. For each region $r = 1,\ldots, b_{num} \times b_{num}$:
  (a) Compute $\hat{A}_U^{H,r}$ of $\hat{A}^{L,r}$ by (14).
  (b) Let $A^{L,r} \leftarrow \hat{A}_U^{H,r}$
  (c) Compute $\hat{A}_V^{H,r}$ of $A^{L,r}$ by (14).

**Return:** Reconstructed result, $A^H \leftarrow \{\hat{A}_V^{H,r}\}_{r=1}^{b_{num}\times b_{num}}$.

## 5. Experimental results

The training database used in the current study consisted of 100 full-frontal facial images of subjects of different ethnicities, all with a neutral expression. The testing database contained 50 frontal-view facial images taken under normal conditions. Subjects in the testing database are not in the training database. Facial images within both databases were manually labeled with seven facial feature points to manually align the facial images. After an affine warp, 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. We use standard SVD to solve (7) and the dimensions of $U$ and $V$ are reduced to retain 98% of the eigenvalue total. The values of $\lambda_U$ and $\lambda_V$ are 0.1 in (11).

Some testing examples are provided (Figure 3(a) and (b)). We compare our algorithm with existing methods, the Cubic B-Spline, Fig. 3(c), and the linear parametric model in [9] (results of [12] are quite similar to results of [9]), Fig. 3(f). To test the performance of 2D formulation; the non-linear method of the local stage in [9] is replaced by its global linear parametric model. The method in [9] is an example of independent approaches, where each image is represented as a 1D vector, and the low-resolution images and the high-resolution images are modeled independently. Results by [9] do not look like the same individual in Fig. 3(b). On the contrary, the results of the proposed method (in Fig. 3(e)) can be recognized as the same individual in Fig. 3(b). Some results of the global stage by 2DDCM approach are displayed in Fig. 3(d). Although the global results are somewhat blurred, they are still clearer than the results of the Cubic B-Spline, Fig. 3(c). Furthermore, reconstructed images in Fig. 3(d) have explicit global geometrical structures, because the $U$ and $V$ space records the significant

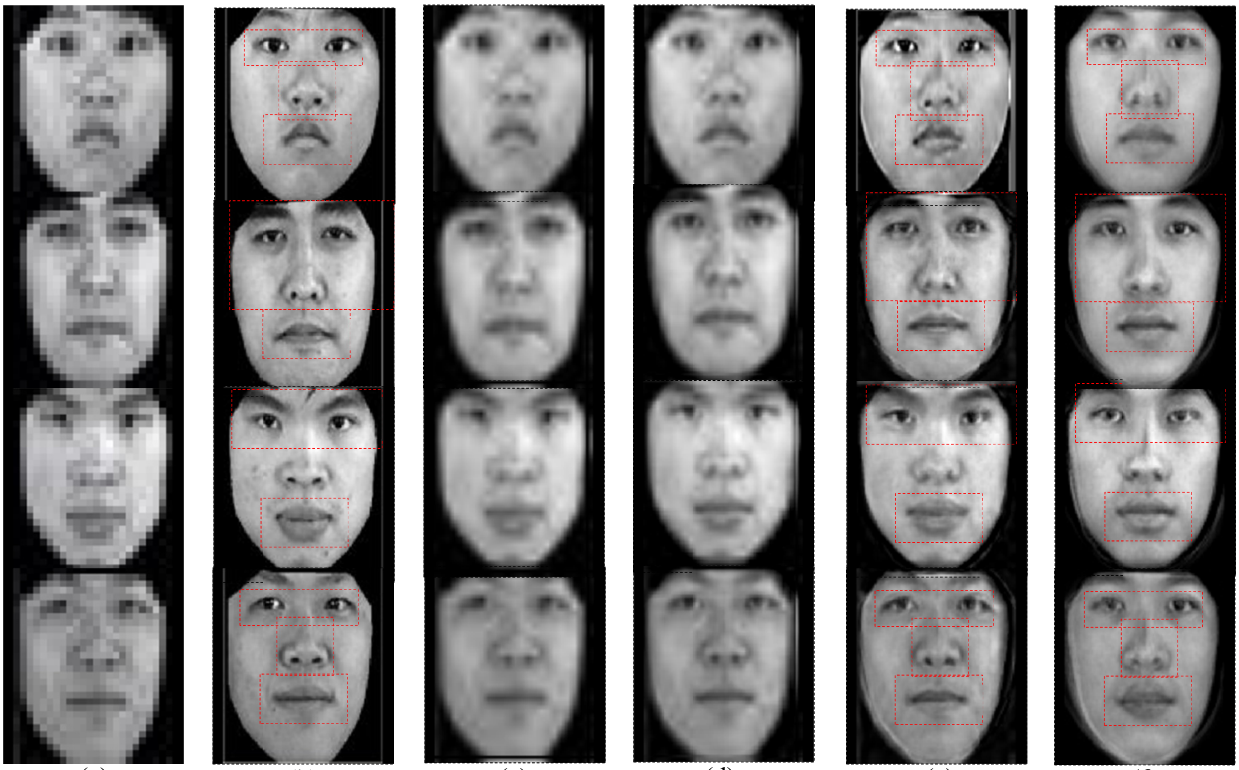


Figure 3. Hallucination results. (a) low-resolution input; (b) original high-resolution image; (c) result by Cubic B-Spline; (d) result of the global stage by the proposed method; (e) final result by the proposed method; (f) result by method in [9]. Note that boxes represent the apparent reconstructed differences between (e) and (f).

characteristics of the vertical and horizontal facial patterns.

## 6. Conclusion

We have described a new method for learning the relationship of two related classes, the low-resolution facial image and its high-resolution version, based on a training dataset. This work was motivated by the inefficiencies and unsatisfactory results of representing two related images by two individual 1D vectors. Compared with existing approaches, the results of 2DDCM algorithm embedding of the vertical and horizontal facial patterns into the column and row spaces is more convenient for improving face hallucination technology.

## References


[1] M. Davis, M. Smith, J. Canny, N. Good, S. King, and R. Janakiraman, "Towards Context-aware Face recognition," *Proc. ACM Multimedia*, pp. 483-486, 2005.

[2] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof, "Fast Active Appearance Model Search Using Canonical Correlation Analysis," *IEEE Trans. on PAMI*, pp. 1690-1694, 2006.

[3] B. Gong, Y. Wang, J. Liu, and X. Tang, "Automatic Facial Expression Recognition on A Single 3D Face by Exploring Shape Deformation," *ACM Multimedia*, pp. 569-572, 2009.

[4] D. Glasner, S. Bagon, and M. Irani. "Super-resolution from a single image," *In Proc. of ICCV*, pp. 349–356 2009.

[5] H. Huang, H. He, X. Fan, and J. Zhang, "Super-resolution of Human Face Image Using Canonical Correlation Analysis," *Pattern Recognition*, Vol. 43, pp. 2532-2543, 2010.

[6] C. Lan, R. Hu, K. Huang, and Z. Han, "Face Hallucination with Shape Parameters Projection Constraint," *ACM Multimedia*, pp. 883-886, 2010.

[7] D. Lin, and X. Tang, "Coupled Space Learning of Image Style Transformation," *Proceedings of the IEEE ICCV*, Vol. 2, pp. 1699-1706, 2005.

[8] Z. C. Lin and H. Y. Shum, "Fundamental Limits of Reconstruction based Super-resolution Algorithms under Local translation," *IEEE Trans. on PAMI*, pp. 83–97, 2004.

[9] C. Liu, H. Shum, and T. Freeman, "Face Hallucination: Theory and Practice," *IJCV*, pp.115-134, 2007.

[10] W. Liu, D. Lin, and X. Tang, "Hallucinating Faces: Tensor-patch Super-resolution and Coupled Residue Compensation," *IEEE CVPR*, pp. 478-484, 2005.

[11] P. Sanguansat, "Face Hallucination Using Bilateral-projection-based Two-dimensional Principal Component Analysis," *International Conference on Date of Conference*, pp. 20-22, 2008.

[12] C. T. Tu and J.J Lien, "Automatic Location of Facial Feature Points and Synthesis of Facial Sketches Using Direct Combined Model," *IEEE Trans. on SMC-b*, Vol. 40, pp. 1158-1169, 2010.

[13] S. Wang, L. Zhang, Y. Liang, Q. Pan, "Semi-Coupled Dictionary Learning with Applications in Image Super-Resolution and Photo-Sketch Image Synthesis," *In Proc. of CVPR*, pp. 2216-2223, 2012.

[14] J. Yang, D. Zhang, A. F. Franji, and J. Y. Yang, "Two Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition, *Trans. PAMI*, Vol.26, pp. 131-137, 200