OCCLUSION INVARIANT FACE HALLUCINATION

Shun-Jen Lee[1,2]*, Meng-Huan Wu*[1,2]*, Chia-Po Wei*[2]*, Yu-Chiang Frank Wang*[2]

[1]Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[2]Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan

ABSTRACT

Face images captured by surveillance cameras are generally noisy and with low resolution (LR). In addition, such images might be corrupted due to occlusion or disguise, which would make the recovery of their high resolution (HR) versions very challenging. In this paper, we propose a face hallucination method which is robust to occlusion or undesirable artifacts. Based on sparse representation, our method is able to identify the corrupted image pixels without any prior knowledge on the type of occlusion for the LR inputs, while the pixels of interest for the HR outputs will be synthesized by solving sparse coding tasks. Experimental results confirm that our approach not only results in satisfactory image quality for the recovered HR outputs, improved recognition performance will also be achieved for LR-to-HR recognition tasks.

Index Terms— Face Hallucination, Face Recognition, Sparse Representation

1. INTRODUCTION

Face hallucination typically refers to the synthesis of a high-resolution (HR) face image from its low-resolution (LR) inputs. Its success would be beneficial for several real-world applications. In this paper, we consider the challenging face hallucination task in which face images might be corrupted due to occlusion. Therefore, our goal is to recover the image pixels of interest in occluded face images.

Given training HR and LR image pairs, most existing face hallucination approaches construct a co-occurrence model for observing the relationships between LR and HR images. Given a test LR image input, the derived model can be directly applied to predict its HR output. Generally, the strategies for performing the above process can be divided into global or local matching based techniques. Global matching based approaches like [1, 2] aim to fit the HR image by the LR ones directly. Despite of computation efficiency, such methods might encounter artifacts in the recovered output and thus require additional post processing processes.

On the other hand, local matching based ones perform the above modeling process at the image patch level. Although improved image quality can be achieved, such methods typically involves more sophisticated learning processes. For example, Chang *et al.* [3] proposed a SR framework based on least squares regression (LSR), which selects candidate patches for determining proper reconstruction weights for input image patches. Ma *et al.* [4] extended the above work by observing the image structure information, and proposed at position-based matching algorithm for face hallucination. Inspired by the above idea, Jung *et al.* [5] advocated the sparse representation of such image patches, while Jiang *et al.* [6] exploited locality information during image synthesis. Recently, coupled-layer information between HR and LR images was further utilized for improved image quality [7].

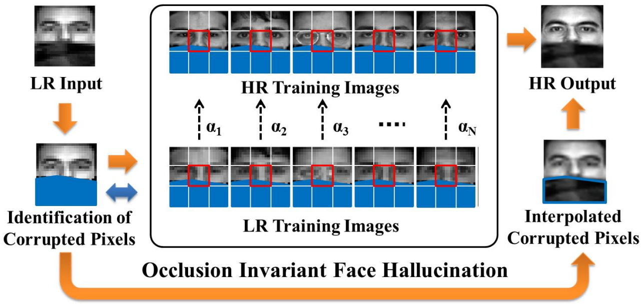
LR Input
HR Training Images
HR Output
α_1
α_2
α_3
....
α_N
Identification of Corrupted Pixels
LR Training Images
Interpolated Corrupted Pixels
Occlusion Invariant Face Hallucination

Fig. 1. Illustration of our proposed framework for occlusion invariant face hallucination. Note that α denotes the coefficient for patch-based image sparse representation.

Nevertheless, the performance of most existing hallucination works would degrade if the input images are corrupted/occluded. To deal with such problems, we propose a sparse representation based scheme for modeling the relationship between HR and LR images in a robust way. To be more specific, our algorithm is able to identify corrupted image pixels during the learning of image representation, which would prevent the degradation of the derived image model by disregarding the poorly reconstructed pixels. As a result, HR images with improved quality will be obtained even if input LR images are corrupted. In our experiments, we will not only show that our method is able to produce satisfactory face hallucination outputs (via both quantitative and qualitative evaluation), we will also confirm that our approach is able to achieve promising recognition results. We will show that our approach would perform favorably against recent approaches on both hallucination and recognition tasks.

2. OUR PROPOSED METHOD

2.1. Occlusion Invariant Sparse Coding

For practical face hallucination scenarios, the input LR images might be corrupted due to occlusion or disguise, and thus it would be crucial to develop a hallucination algorithm which is robust to such undesirable effects. In our work, we propose a sparse representation based algorithm to recover the image regions of interest in the HR outputs, while the corrupted ones will be simplified predicted by interpolation based techniques (and be disregarded if recognition is performed).

Before detailing our proposed method, we first define the notations and variables for the sake of clarification. Let $\mathbf{y}_L \in \mathbb{R}^d$ be the LR input image, which might be partially occluded. We have $\mathbf{D}_L \in \mathbb{R}^{d\times n}$ as the LR training dictionary of n instances, i.e., $\mathbf{D}_L = [\mathbf{x}_{L,1}, \mathbf{x}_{L,2}, \cdots, \mathbf{x}_{L,N}]$, where $\mathbf{x}_{L,i} \in \mathbb{R}^d$ denotes the ith LR training image and N is the number of training images. To solve image representation problems with occlusion, Yang *et al.* [8] recently presented robust sparse coding (RSC), which represents the LR input by solving

$$\min_{\boldsymbol{\alpha}_L} \|\mathbf{W}_L(\mathbf{y}_L - \mathbf{D}_L\boldsymbol{\alpha}_L)\|_2^2 + \lambda \|\boldsymbol{\alpha}_L\|_1, \tag{1}$$

where the weight matrix $\mathbf{W}_L$ is defined as

$$\begin{aligned} &\mathbf{W}_L = \text{diag}(w(e_1), w(e_2), \ldots, w(e_d))^{1/2}, \\ &\text{and} \;\; w(e_k) = \frac{exp(-\mu e_k^2 + \mu\delta)}{1 + exp(-\mu e_k^2 + \mu\delta)}. \end{aligned} \tag{2}$$

In (2), e_k indicates the kth entry of $\mathbf{e} = \mathbf{y}_L - \mathbf{D}_L\boldsymbol{\alpha}_L$, which represents the reconstruction error for the kth image pixel in the LR input. The function $w(e_k)$ is designed to produce a small value when the magnitude of e_k is large (and vice versa). In [8], the product $\mu\delta$ is chosen to be a large constant, and δ is the jth largest entry in the vector $[e_1^2, e_2^2 \ldots e_d^2]$. RSC has j as the nearest integer to τd with $\tau \in [0.6, 0.8]$. Nevertheless, the goal of RSC is to suppress the influence of poorly reconstructed pixels.

It is worth noting that τ can be interpreted as the proportion of non-occluded image pixels to the total pixel number in the LR input. For example, if we set τ as 0.6, 40% percent of pixels associated with the largest (poorest) reconstruction errors will be suppressed. In other words, the remaining 60% of the pixels will be recovered by $\mathbf{D}_L$. Unfortunately, there is no guideline in [8] to determine this crucial parameter τ, and thus it might not be easy to apply RSC for practical scenarios.

To overcome the above problem, we propose a maximum likelihood estimation (MLE) [9] based approach for determine the optimal τ, which allows us to recover the non-occluded image regions with performance guarantees. Instead of setting τ as a fixed value for all input images, we first observe the recovered images with varying τ values. Figure 2 shows an example, in which each point denotes the total reconstruction error for non-occluded image pixels with a particular τ value. More specifically, by varying the value of τ, we obtain a set of points $(\tau_1, e_1), \ldots, (\tau_M, e_M)$.

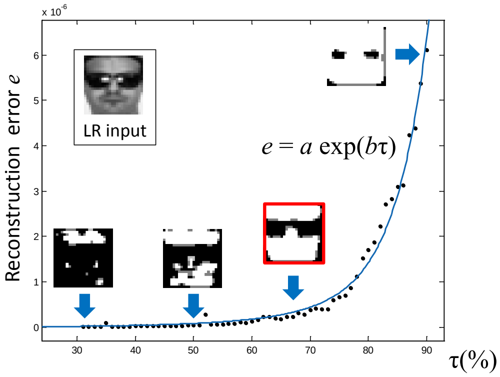

Fig. 2. Reconstruction errors of non-occluded image regions with varying τ values. The blue curve is an exponential function $e = a\exp(b\tau)$ fitting the observed instances. Note the the preferable recovered image is bounded by a red rectangle.

In order to determine the optimal τ for obtaining the preferable recovered face image (i.e., the one with red rectangle in Figure 2), we need to first fit the observed points with an exponential function, i.e.,

$$e = f(\tau, a, b) = a\exp(b\tau), \tag{3}$$

where a and b are the parameters to be determined. Note that $f(\tau, a, b)$ in (3) is a nonlinear function of τ. To make the fitting problem more tractable, we take the logarithm of both sides of (3) and obtain $\ln(e) = \ln(a) + b\tau$. Based on MLE, we approach this curve fitting task by solving the following optimization problem:

$$\min_{a,b} \sum_{m=1}^{M} \left(\ln(e_m) - \ln(a) - b\tau_m\right)^2. \tag{4}$$

By defining $\tilde{a} = \ln(a)$, the above minimization problem can be expressed as

$$\min_{\tilde{a},b} \left\| \tilde{\mathbf{e}} - [\mathbf{1}, \boldsymbol{\tau}] \begin{bmatrix} \tilde{a} \\ b \end{bmatrix} \right\|_2^2, \tag{5}$$

where $\tilde{\mathbf{e}} = [\ln(e_1), \ldots, \ln(e_M)]^T$, $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_M]^T$, and $\mathbf{1} = [1, \ldots, 1]^T \in \mathbb{R}^M$. We note that (5) is a standard least squares problem, and its analytical solution is dervied by

$$\begin{bmatrix} \tilde{a} \\ b \end{bmatrix} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\tilde{\mathbf{e}} \tag{6}$$

with $\boldsymbol{\Phi} := [\mathbf{1}, \boldsymbol{\tau}]$. Once we have $\tilde{a}$, the parameter a is obtained as $a = \exp(\tilde{a})$.

With parameters a and b for the exponential curve (3) learned by MLE, we now discuss how to determine the optimal τ for recovering the image without occluded pixels (e.g., the one bounded by the red rectangle in Figure 2). In our

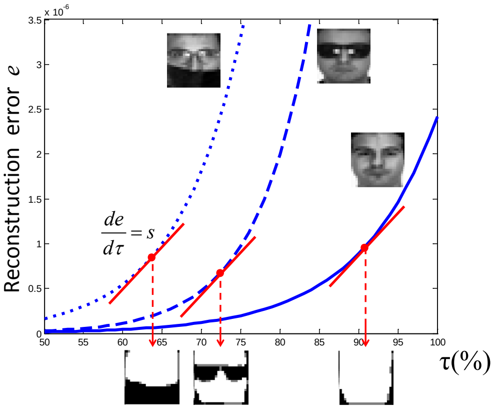

Fig. 3. Fitting curves of reconstruction errors e vs. τ for three face images with different degrees of occlusion. The points of tangency depict the optimal τ for recovering non-occluded image regions.

work, we empirically observe that the optimal τ always corresponds to the intersection of the reconstruction curve and the straight line with a fixed slope s as shown in Figure 3. Moreover, this slope s does *not* vary with different types of face images (occluded or not).

In our work, the value of s can be directly observed from training data, which is a constant over images with different degrees of occlusion (as illustrated in Figure 3). Once the value of s is known, we can begin to calculate parameter τ. To solve this task, we set the derivative of (3) with respect to τ equal to s. As a result, we have $\frac{de}{d\tau} = ab\exp(b\tau) = s$, and the optimal τ is obtained as follows

$$\tau = \frac{1}{b}(\ln(s) - \ln(ab)). \tag{7}$$

Different from RSC which requires one to manually select τ for representing the input image, our method is able to automatically select the best τ for each input without any prior knowledge on the type of occlusion, which is preferable for practical hallucination tasks. Thus, we refer to this proposed encoding process as *occlusion invariant sparse coding*.

Once the optimal τ, we apply (2) for calculating the associated weight matrix $\mathbf{W}_L$. For refinement purposes, a median filter is applied to the resulting $\mathbf{W}_L$ which preserves the completeness and smoothness of the determined occluded regions.

2.2. Occlusion Invariant Sparse Representation for Face Hallucination

With the proposed occlusion invariant sparse coding scheme in Section 2.1, we are able to derive the sparse coefficients for representing the LR input $\mathbf{y}_L$ using LR dictionary $\mathbf{D}_L$. To predict the corresponding HR output image, we apply the sparse representation based super resolution techniques of [10, 5] for our face hallucination task, as we now discuss.

Given the LR dictionary and its corresponding HR version $\mathbf{D}_H \in \mathbb{R}^{m\times n}$, it is assumed in [10, 5] that the sparse coefficients for representing the HR image $\mathbf{y}_H \in \mathbb{R}^m$ would be identical to those for its LR input $\mathbf{y}_L \in \mathbb{R}^d$. However, since approaches like [10, 5] are not robust to occluded regions, we need to apply the proposed occlusion invariant sparse coding scheme for addressing this challenging problem.

Recall that, in Section 2.1, our proposed algorithm is able to derive the weighting matrix $\mathbf{W}_L$ from $\mathbf{y}_L$ and $\mathbf{D}_L$, which indicates and recovers the non-occluded image regions while suppressing corrupted pixels. For hallucinating the HR image, we first extend the derived matrix $\mathbf{W}_L$ to $\mathbf{W}_L \in \mathbb{R}^{m\times m}$ via interpolation. Next, we consider patch-based sparse representation for performing occlusion invariant face hallucination. That is, we divide the LR input $\mathbf{y}_L$ into overlapping patches, denoted by $\mathbf{y}_L^i$, where i is the patch index. Similary, each LR training image in $\mathbf{D}_L$ is also divided into overlapping patches, denoted by $\mathbf{x}_{L,j}^i$ for $j = 1, 2, \ \ldots, N$.

Now, the sparse coefficient for the ith patch can be determined by solving the following problem:

$$\min_{\boldsymbol{\alpha}_L^i} \left\|\mathbf{W}_L^i(\mathbf{y}_L^i - \mathbf{D}_L^i\boldsymbol{\alpha}_L^i)\right\|_2^2 + \lambda \left\|\boldsymbol{\alpha}_L^i\right\|_1, \tag{8}$$

where $\mathbf{W}_L^i$ indicates the ith patch of $\mathbf{W}_L$. With $\boldsymbol{\alpha}_L^i$ calculated, the ith output patch can be derived by

$$\mathbf{y}_H^i = \mathbf{W}_H^i\mathbf{D}_H^i\boldsymbol{\alpha}_L^i, \tag{9}$$

where $\mathbf{W}_H^i$ is the ith patch in $\mathbf{W}_H$, and $\mathbf{D}_H^i$ indicates the ith patches of $\mathbf{D}_H$. Once all the patches for the HR output is obtained, hallucination of $\mathbf{y}_H$ is complete. We note that, for occluded image regions in $\mathbf{y}_H$, we simply perform bicubic interpolation on the associated regions of the LR input. The framework of our method is depicted in Figure 1.

3. EXPERIMENTS

3.1. Database and Settings

We consider the AR database [11] for evaluation, which contains 126 individuals with more than 4000 frontal face images. In our experiment, we consider a subset of AR by randomly choosing 100 individuals with 50 men and 50 women. All images are converted into grayscale. We crop and rescale the images into 96×96 and 24×24 pixels as HR and LR images, respectively. For each subject in AR, we only consider the neutral image and images with expression variations from Session 1 in the training set. The remaining ones in both sessions are viewed as test images.

3.2. Face Hallucination

To evaluate our face hallucination results, we consider the LR face image inputs with and without occlusion. In addition, since our method is able to identify the corrupted pixels automatically, we further include the evaluation of recovered HR images using non-occluded pixels only. This is to show that,

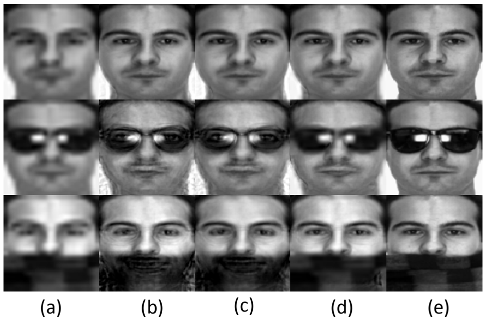

Fig. 4. Example face hallucination results including occluded image regions. Images in each row are outputs of (a) Bicubic, (b) Jung *et al.* [5], (c) Jiang *et al.* [6], (d) ours, and (e) the ground truth.

Table 1. Comparisons of average PSNR & SSIM values of face hallucination outputs with occlusion.

	Bicubic	[5]	[6]	Ours
PSNR	24.09	23.67	24.05	**24.58**
SSIM	0.7975	0.7240	0.7404	**0.8094**

after disregarding such undesirable pixels, our method is able to achieve improved image quality for face hallucination.

We first compare the example outputs in Figure 4, in which the HR images (including the occluded image regions) are produced by bicubic, the approaches of [5] and [6], and ours. In addition, we also show the ground truth HR images in the last column of Figure 4. From this figure, we see that our approach achieved satisfactory image quality for non-occluded image regions. Recall that our method views occluded regions as image corruption, and thus the corresponding pixels are produced by bicubic interpolation. In Table 1, we quantitatively compare the image quality using PSNR and SSIM values. It can be seen that our approach was able to achieve improved results than other SR methods did.

Next, we consider the comparisons in which the HR images are corruption free. In other words, we apply our proposed method to remove the undesirable image pixels (mainly due to occlusion) from the HR images of different approaches. Figure 5 and Table 2 present and compare the results of different approaches. We see that, after removing such corrupted image pixels, our method still obtained improved image quality for the remaining pixels of interest, and thus performed favorably against other SR approaches. Therefore, from the above qualitative and quantitative evaluation, the effectiveness of our proposed method for occlusion invariant face hallucination can be successfully verified.

3.3. LR-to-HR Face Recognition

We now consider face recognition with LR images as test inputs, while *only* the HR ones for training. This is to confirm that, in addition to improved HR outputs, our proposed method can be applied to LR-to-HR face recognition.

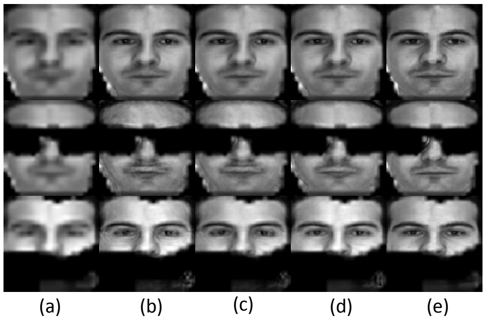

Fig. 5. Example face hallucination results with corrupted pixels removed. Images in each row are outputs of (a) Bicubic, (b) Jung *et al.* [5], (c) Jiang *et al.* [6], (d) ours, and (e) the ground truth.

Table 2. Comparisons of average PSNR & SSIM values of face hallucination outputs without corrupted pixels.

	Bicubic	[5]	[6]	Ours
PSNR	26.40	26.64	27.05	**27.13**
SSIM	0.8659	0.8588	0.8672	**0.8806**

Table 3. LR-to-HR recognition performance on AR.

LR/LR	Bicubic	[5]	[6]	Ours	HR/HR
82.21	80.33	79.68	76.63	87.37	87.79

To recognize the LR images, we first apply our method to recover their HR versions. Next, the approach of sparse representation based classification (SRC) [12] will be applied to perform recognition. We note that, for comparison purposes, LR/LR (HR/HR) in Table 3 indicates the direct use of LR (HR) images for both training and testing. Their performances can be viewed as lower/upper bounds for recognition. In addition to bicubic interpolation, we also consider recent approaches of [5] and [6] to synthesize the HR outputs for recognition. Table 3 compares the recognition results of different methods. From this table, we see that our approach achieved the highest recognition rate, which supports the the use of our proposed scheme for LR-to-HR face recognition.

4. CONCLUSION

We presented a sparse representation based approach for occlusion invariant face hallucination. Our proposed scheme extends robust sparse coding for determining the optimal parameter automatically. As a result, non-occluded image regions can be recovered from the LR inputs, while their corresponding HR versions can be synthesized accordingly. The above coding and learning process is based on maximum likelihood estimation, and thus no user interaction or parameter tuning is required. Experimental results on both face hallucination and the LR-to-HR face recognition tasks confirmed the effectiveness and robustness of our approach.

5. REFERENCES

[1] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2005.

[2] J.-S. Park and S.-W. Lee, "An example-based face hallucination method for single-frame, low-resolution facial images," *IEEE Trans. on Image Processing*, 2008.

[3] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE CVPR*, 2004.

[4] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognition*, 2010.

[5] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *Signal Processing Letters*, 2011.

[6] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. on Multimedia*, 2014.

[7] J. Jiang, R. Hu, L. Chen, Z. Han, T. Lu, and J. Chen, "Coupled-layer neighbor embedding for surveillance face hallucination.," in *Proc. IEEE ICIP*, 2013.

[8] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proc. IEEE CVPR*, 2011.

[9] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[10] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. on Image Processing*, 2010.

[11] A. M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, 1998.

[12] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009.