In this paper, we formulate the face hallucination as an image decomposition problem, and propose a Morphological Component Analysis (MCA) based method for hallucinating a single face image. A novel three-step framework is presented for the proposed method. Firstly, a low-resolution input image is up-sampled via an interpolation. Then, the interpolated image is decomposed into a global high-resolution image and an unsharp mask by using MCA. Finally, a residue compensation is performed on the global face to enhance its visual quality. In our proposal, the MCA plays a vital role as MCA can properly decompose a signal into several semantic sub-signals in accordance with specific dictionaries. By virtue of the multi-channel decomposition capability of MCA, the proposed method can be also extended to simultaneous implementation of face hallucination and expression normalization. Experimental results demonstrate the effectiveness of our method for the images from both lab environment and realistic scenarios. We also study the contribution of face hallucination to face recognition in the case that probe images and gallery images are under different resolutions. The main conclusion is that the contribution is significant when using local facial features (e.g., LBP), but unobvious when using holistic facial features (e.g., Eigenfaces).

© 2012 Elsevier B.V. All rights reserved.
LR face and the learned pixel structures as priors to estimate the target HR face. Yang et al. [32] proposed an image super-resolution method via sparse representation. In their method, a nonnegative matrix factorization (NMF) is used for zooming the LR face image to a medium resolution, and then a sparse representation is applied on low dimensional image patches for recovering local details.

Although the existing face hallucination methods have achieved great success, there is still large room for improvement. One need to improve is the estimation of global face because existing algorithms often result in a “fake hallucination” especially causing ringing artifacts. One of the reasons is that, only global face rather than all components of input image is modeled in existing estimation models, such may reduce the accuracy of estimation. This paper therefore focus on making the global face estimation better and comes up with a new face hallucination method.

1.2. Proposed method

The proposed method is under a three-step framework, where the face hallucination is formulated as a novel image decomposition problem. The recently proposed morphological component analysis (MCA) [8–11] is used for such a decomposition. At the first step, the LR input is up-sampled by an interpolation. We observe that the interpolated image can be represented as a superposition of two typical components, i.e., a global HR image and an “unsharp mask”. In the second step, we use MCA to implement such a decomposition so as to obtain the global approximation of the HR image from interpolated image. In the third step, facial detail information is compensated onto the estimated HR image by using the neighbor reconstruction of position-patches. Our preliminary work was first reported in [12].

Owing to the typical multi-channel decomposition capability of MCA, we show that our method can be extended to perform face hallucination and expression normalization at the same time, where a special dictionary is designed to model the expression variation in MCA. To the best of our knowledge, this is the first paper to study the contribution of face hallucination to face recognition. Conclusions are given in Section 4.

2. Proposed methodology

2.1. Problem formulation

A classical resolution reduction model for single image can be represented as [13]

\[ x_l = s \downarrow (h^* x_h) \]  

(1)

which explains the relationship between the observed LR images \( x_l \) and the ground truth HR image \( x_h \). \( s \downarrow \) represents the down-sampling operator by a factor of \( s \). \( h \) is the camera’s point spread function (PSF) acting as a blurring operator, and * denotes the convolution operation. Eq. (1) is usually called reconstruction constraint of image super-resolution.

Now we apply an up-sampling operator \( s \uparrow \) onto (1):

\[ x_{zoom} = s \uparrow x_l = s \uparrow s \downarrow (h^* x_h) \approx h^* x_h, \]  

(2)

where \( x_{zoom} \) denotes the HR image interpolated from \( x_l \). We use a bilinear interpolation \( s \uparrow \) to simulate the inverse of \( s \downarrow \), with the fact that no an actual inverse may be discovered. In this case, \( x_{zoom} \) can be regarded as an approximate blurred version of \( x_h \) that can be formulated as follows according to Gonzalez and Woods [14]:

\[ x_{zoom} = x_h + x_l, \]  

(3)

where \( x_h \) is a high-frequency component containing both positive and negative values, termed “unsharp mask”. Based on (3), the critical step of face hallucination in our formulation is how to separate \( x_h \) and \( x_l \) from \( x_{zoom} \). In the next section, we investigate how to employ the MCA for this task.

2.2. Face image decomposition by MCA

2.2.1. A brief overview of MCA

In a series of recent literatures [8–11], the MCA was developed for separating a signal into several components which have different morphologies. The main idea behind MCA is to use the morphological diversity of the different features contained in the data, e.g., for our application, HR face image \( x_h \) and unsharp mask \( x_l \) in interpolated image \( x_{zoom} \), and to associate each morphology with a dictionary of atoms [11]. After sparsifying the original signal with different dictionaries, a plausible separation of sub-signals can be generated. Unlike PCA which models
only one type of content of the signal (e.g., piecewise smooth content of an image), MCA can simultaneously model all sub-bands, such that it can properly decompose a signal into several semantic parts.

### 2.2.2. Usage of MCA

Towards the problem formulated in (3), we have the following optimization function based on MCA:

$$
\min_{\mathbf{a}_h, \mathbf{a}_s} \left( \left| \left| \mathbf{z}_h \right| \right|_0 + \left| \left| \mathbf{z}_s \right| \right|_0 \right),
$$

s. t. \quad \left| \left| \mathbf{X}_{zoom} - \mathbf{P}_h \mathbf{a}_h - \mathbf{P}_s \mathbf{a}_s \right| \right|_2 \leq \epsilon

(4)

where $\epsilon$ is an error tolerance, which in practice is potentially determined by keeping 99% energy in PCA (see Section 3.1.1 for more details). $\mathbf{P}_h$ and $\mathbf{P}_s$ are the dictionaries associated with $\mathbf{z}_h$ and $\mathbf{z}_s$, respectively. The solution $(\mathbf{z}_h, \mathbf{z}_s)$ of (4) would generate a sub-signals separation with $\mathbf{z}_h = \mathbf{P}_h \mathbf{a}_h$ and $\mathbf{z}_s = \mathbf{P}_s \mathbf{a}_s$.

A fundamental problem to apply MCA is the design of dictionaries. Although pre-constructed dictionaries like wavelets [15], contourlets [16], and curvelets [17] have been successfully used in image decomposition, they are effectual for general image reconstruction but not necessarily good for specific decomposition task. In this paper we adopt the PCA to adaptively learn the dictionaries, since PCA can reconstruct certain high-dimension signals by using very few principal components. Given training samples of HR face images and unsharp masks by PCA, we obtain their mean values $\mathbf{x}_h$ and $\mathbf{x}_s$, all eigenvalues $(\sigma_k(i))_{i=1}^N$ and $(\sigma_k(i))_{i=1}^N$, and corresponding eigenvector matrices $\mathbf{P}_h$ and $\mathbf{P}_s$, respectively. Accordingly, any $\mathbf{x}_h$ and $\mathbf{x}_s$ can be respectively represented as

$$
\mathbf{x}_h \approx \mathbf{x}_h + \mathbf{P}_h \mathbf{z}_h, \quad \mathbf{x}_s \approx \mathbf{x}_s + \mathbf{P}_s \mathbf{z}_s,
$$

(5)

where $\mathbf{z}_h$ and $\mathbf{z}_s$ are projection coefficients. Then $\mathbf{P}_h$ and $\mathbf{P}_s$ can be used as the dictionaries for describing $\mathbf{x}_h - \mathbf{x}_h$ and $\mathbf{x}_s - \mathbf{x}_s$, respectively.

Essentially, MCA is to excavate the morphological diversity of different sub-signals, adoptive dictionaries should be as incoherent as possible between each other. Now we investigate the coherence of dictionaries learned by PCA in our issue. The following mutual coherence [10], which is a measurement of dependence between the atoms of two dictionaries $\mathbf{P}_h$ and $\mathbf{P}_s$, has been considered:

$$
\mu(\mathbf{P}_h, \mathbf{P}_s) = \max_{i,j} \left| \phi^*_{h,i} \phi_{s,j} \right|,
$$

(6)

where $\phi^*_{h,i} \phi_{s,j}$ denotes the entry in the $i$th row and the $j$th column of inner product $\phi^*_{h,i} \phi_{s,j}$. Assuming the columns of $\mathbf{P}_h$ and $\mathbf{P}_s$ are normalized to unit l2-norm, the value range of mutual coherence is [0,1], where 0 means the atoms from two dictionaries are mutually orthogonal and 1 means there are same atoms from different dictionaries. Therefore, a smaller one stands for more incoherence between two dictionaries. In our practice, the designed dictionaries with mutual coherence smaller than 0.5 are accepted. In our experiment, the mutual coherence of $\mathbf{P}_h$ and $\mathbf{P}_s$ is 0.3303, which is small enough to ensure the incoherence.

When using the dictionaries learned by PCA for problem (4), two constraints should be considered. The first is to constrain coefficients $\mathbf{z}_h$ and $\mathbf{z}_s$ within an appropriate range so as to produce a rational reconstruction. Accordingly, we restrict $\mathbf{z}_h$ and $\mathbf{z}_s$ within a hyper-ellipsoid $[-3\sigma(i), 3\sigma(i)], \sigma = h, s$. The second one is an assumption of following relationship between the LR input $\mathbf{x}_i$ and the smoothed down-sampled HR image $\mathbf{x}_h$ [3,13]

$$
\mathbf{x}_i(m,n) = \frac{1}{S^2} \sum_{j=1}^S \sum_{i=1}^S \mathbf{x}_i(Sm + i, Sn + j),
$$

(7)

where $(m,n)$ denotes the coordinate of pixel, and $S$ denotes the magnification factor. In this paper, we use the common setting $S = 4$ or 8. Based on above analysis, (4) is specified as

$$
\min_{\mathbf{z}_h, \mathbf{z}_s} \left( \left| \left| \mathbf{z}_h \right| \right|_0 + \left| \left| \mathbf{z}_s \right| \right|_0 \right),
$$

s. t. \quad \left| \left| \mathbf{X}_{zoom} - \mathbf{X}_{h} - \mathbf{P}_h \mathbf{z}_h - \mathbf{P}_s \mathbf{z}_s \right| \right|_2 \leq \epsilon,

$$
-3\sigma(i) \leq \mathbf{z}_h(i) \leq 3\sigma(i), \quad 3 = h, s, \quad i = 1, \ldots, N,

$$
\mathbf{x}_i(m,n) = \frac{1}{S^2} \sum_{j=1}^S \sum_{i=1}^S \mathbf{x}_i(Sm + i, Sn + j).
$$

(8)
Although Eq. (8) is NP-hard in general, under certain conditions, the sparsest solution is also the solution with minimal l1-norm [18]. In practice, it can be effectively solved in an alternately iterative thresholding manner [19]. At each iteration step, the threshold results are rescaled into the allowed hyper-ellipsoid range, and a back-projection method [20] is used to correct the generated HR image part to meet the reconstruction constraint. The proposed decomposition method using MCA is summarized in Table 1. Although there is no way to optimally estimate the thresholds and the number of iterations yielding a successful separation, we can use experience to achieve satisfying results. The detailed setting can be seen in Section 3.1.1.

In this MCA-based method, because HR image and unsharp mask are modeled simultaneously in a framework, a more rational separation than single-channel method can be obtained. It should be noted that in our method, PCA is used for obtaining an initial estimation of the HR face rather than for getting the ultimate super-resolution result. Based on the rationale of MCA [8–11], the PCA with l0 penalty can automatically choose the most efficient principal components to model each content type, enabling a successful separation of image contents for MCA. In other words, the PCA with l0 penalty achieves a “special” modeling but the original PCA results in a more “common” modeling for the content type. In our proposed method, by enforcing the l0 penalty, multiple content types such as the high-resolution image and the unsharp mask can be appropriately separated from an interpolated face. The estimation result in Fig. 1 verifies that the PCA with l0 penalty works much better than that with l2 penalty.

To enhance the visual quality, it is necessary to compensate local details onto the estimated global HR face image. In the next step, we will introduce a residue compensation algorithm for this task.

### 2.3. Residue compensation

In this paper, we use the residue compensation [5] to enrich the facial details of HR image. In this section, we re-note the global HR face image which has been estimated by MCA as $\mathbf{x}_g^h$. Furthermore, define a LR residue as follows:

$$
\mathbf{x}_l^l = \mathbf{x}_l - \frac{1}{S^2} \sum_{i=1}^{S} \sum_{j=1}^{S} \mathbf{e}_l^i (S m + i, S n + j).
$$

(9)

Then, we need to estimate a HR residue $\mathbf{x}_l^h$ from $\mathbf{x}_l^l$. We select the neighbor reconstruction of position-patches (NRPP) [6] for this task. The basic idea of NRPP is to regard each residue image as a patch matrix composed of overlapped square patches. For a given LR residue patch, it can be reconstructed by a combination of k-nearest neighboring LR training residue patches at the same facial position. Then, we can synthesize the corresponding HR residue patch by replacing the LR training residue patches
with the corresponding HR training residue patches, while maintaining the same combination weights. Finally, the HR residue face $x_l^h$ is obtained by integrating these HR residue patches by averaging the overlapped parts. The derived HR residue face $x_l^h$ is then added to the global face $x_g^h$ to obtain the final hallucinated face:

$$x_h = x_l^h + x_g^h$$

Fig. 2 shows the diagram of the proposed face hallucination method.

### 2.4. An extension: simultaneous super-resolution and expression normalization

In practical applications, e.g., video surveillance, the captured face images are always in low resolution and with various expressions. In this case, simultaneous super-resolution and expression normalization is necessary for improving the performance of face recognition. However, according to our knowledge, such a challenging task is almost impossible to complete by existing methods. Thanks to the multi-channel decomposition capability of MCA, our method can be extended to achieve such a goal, i.e., obtaining a HR face image with neutral expression from an LR input with various expressions, which is described as follows.

Note an observed LR expression image as $x_l$, and its interpolated HR image as $x_l^{zoom}$. Suppose $x_l^{zoom}$ to be a superposition of two components: the interpolated HR image with neutral expression $x_n^{zoom}$ and an “expression mask” $x_e$, that is

$$x_l^{zoom} = x_n^{zoom} + x_e$$

According to Eqs. (3) and (11), we have a new decomposition model:

$$x_l^{zoom} = x_h + x_s + x_e,$$  \hspace{1cm} (12)

where $x_h$ is the HR image with neutral expression, and is what we want to obtain. Intuitively, Eq. (12) can be also solved by MCA with three dictionaries, one of which is for especially modeling $x_e$. However, it is difficult to learn a common dictionary for modeling various expressions, since the expression variations of different persons fall in multiple linear subspaces. In this paper, we adopt a “divide and conquer” manner for expressions modeling, that is, learning five dictionaries \{Pe\} by PCA corresponding to five expressions: close eyes, frown, smile, surprise and open mouth, respectively. In our experiment, the mutual coherences between $P_h$ and five $P_e$ are 0.3374, 0.3483, 0.3043, 0.3173, and 0.3101, respectively. The ones between $P_s$ and five $P_e$ are 0.1273, 0.1422, 0.1327, 0.1332, and 0.1394, respectively. These mutual coherences are small enough for employing MCA. For an input LR face image $x_l$, we determine its expression by using Local Binary Patterns (LBP) based classification [21]. If it is neutral, we do the MCA-based decomposition directly based on (3), otherwise, based on (12) with appropriate expression dictionary.

It should be noted that, beside the number of dictionaries, there are two main differences between the super-resolution algorithm based on (12) and the one based on (3). Firstly, the back-projection correction in the decomposition process cannot be used in the algorithm based on (12). Secondly, in the residue compensation step, for the input LR expression image $x_l$, we need an additional operation to decompose $x_l$ into a neutral LR image $x_l^n$.

Fig. 3. Diagram of the extended method for simultaneous super-resolution and expression normalization, where the images are calibrated for display.
and a LR expression mask $x_{le}$:

$$x_l = x_l^0 + x_{le}, \quad (13)$$

which also can be solved by MCA. Then, the LR residue is computed by

$$x_l^0 = x_l - \frac{1}{S^2} \sum_{i=1}^{S} \sum_{j=1}^{S} x_l^S (S_m + i, S_n + j).$$

Fig. 3 shows the diagram of the extended method for simultaneous super-resolution and expression normalization. In the next section, we will show the experimental results with respect to visual quality and face recognition.

3. Experimental results and discussion

In this section, we conduct experiments to evaluate the proposed face hallucination method mainly in terms of visual quality, and then study the contributions of face hallucination for face recognition. As a comparison purpose, the bilinear interpolation and six state-of-the-art methods including Baker and Kanade’s [3], Wang and Tang’s [4], Liu et al.’s [5], Ma et al.’s [6], Hu et al.’s [7], and Yang et al.’s [32] were also implemented. The experiments were conducted on the CAS-PEAL-R1 face database [22], which consists of 99,594 images of 1040 individuals. All face images were cropped to the size of 128 × 96 pixels, with simple alignment according to the locations of eyes and mouths. As shown in Table 2, five subsets of frontal face images were used in the following three experiments.

3.1. Evaluation by visual quality

3.1.1. Experimental setting

In this experiment, the Normal subset of CAS-PEAL-R1, in which the face images are under normal condition, i.e., frontal pose, even illumination, and neutral expression were used. The Normal subset contains 1040 images for one image per person. By smoothing and down-sampling the original HR images with magnification factor $S=4$ (or $S=8$), the corresponding LR images with size of 32 × 24 (or 16 × 12) pixels were obtained. We call one HR image and its corresponding LR image as a HR–LR image pair. For each magnification factor $S$, we chose 1000 h-LR image pairs as training set and the remaining 40 image pairs as testing set. Specially, for the methods under “global face+local face” framework, 700 image pairs were used for training the global model, and 300 ones were used for training the residue compensation model.

<table>
<thead>
<tr>
<th>Subset</th>
<th>#Variations</th>
<th>#Individuals</th>
<th>#Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1</td>
<td>1040</td>
<td>1040</td>
</tr>
<tr>
<td>Expression</td>
<td>5</td>
<td>377</td>
<td>1884</td>
</tr>
<tr>
<td>Background</td>
<td>2–4</td>
<td>297</td>
<td>651</td>
</tr>
<tr>
<td>Aging</td>
<td>1</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Lighting</td>
<td>≥ 9</td>
<td>233</td>
<td>2450</td>
</tr>
</tbody>
</table>

Fig. 4. Hallucinated results of different methods with magnification factor 4 (from 32 × 24 to 128 × 96): (a) LR input (32 × 24); (b) bilinear interpolation; (c) Baker and Kanade’s [3]; (d) Wang and Tang’s [4]; (e) Liu et al.’s [5]; (f) Ma et al.’s [6]; (g) Hu et al.’s [7]; (h) Yang et al.’s [32]; (i) Proposed method; and (j) ground truth HR images (128 × 96).
For the MCA parameters, the maximum thresholds were automatically set by using the maximum absolute value of all coefficients, i.e., \[ \text{max} T_h = \| P_h (x_{zoom} - x_0 - x_s) \|_\infty \]
and \( \max T_s = \| P_s^T (x_{zoom} - x_n - x_s) \|_{\infty} \). Each minimum stop threshold was set as the square root of the \( t \)th eigenvalue, ensuring that 99% of the variance was captured by the \( t \) principal components corresponding to the largest \( t \) eigenvalues. We observed that when the number of iteration was larger than 50, our algorithm can obtain a stable output. For the back-projection operator, the bi-cubic interpolation was employed.

### 3.1.2. Results

Figs. 4 and 5 illustrate the results of face hallucination with 4-time magnification (from \( 32 \times 24 \) to \( 128 \times 96 \) pixels), and 8-time magnification (from \( 16 \times 12 \) to \( 128 \times 96 \) pixels) by using different methods, respectively. As shown, all other methods got sharper and clearer results than what bilinear interpolation method attained. However, Baker and Kanade’s method may distort some facial details. Wang and Tang’s method works well but may cause ringing effect, especially on face outlines. Hu et al.’s results are relatively blurry, even though they are sharper than that by bilinear interpolation. Liu et al.’s and Yang et al.’s results have some obvious artifacts for 8-time super-resolution. In comparison, it is pretty hard to determine which method works better among Ma et al.’s and ours through the visual inspection. Therefore, to quantify the image visual quality, we also calculated two indexes to measure the difference between hallucinated face and ground truth HR face. The first one is the Mean Square
Error (MSE) which is defined by

$$MSE = \sum_{i=1}^{N} \frac{\left\| h_{\text{test},i} - j_{\text{test},i} \right\|^2}{(W \times H \times N)},$$

where $h_{\text{test},i}$ is the ground truth HR face, $j_{\text{test},i}$ is the hallucinated face from LR face. $W$ and $H$ are the width and height of the image, respectively, and $N$ is the number of test images. The other index is the structural similarity (SSIM) [23], which is under the assumption that human visual perception is highly adapted for extracting structural information from a scene. A lower (higher) value of MSE (SSIM) stands for a better image quality of the hallucinated face. The MSE and SSIM of each method is shown in Figs. 6 and 7, respectively. They show that our method gets the best image quality evaluation, except for comparing with Ma et al.’s method for 4-time super-resolution case. We believe the main reason of making our result better is the good estimation of global face by MCA.

In order to investigate the performance of the global face modeling step in the proposed method, we illustrate in Fig. 8 the comparison results of the global faces produced by different global methods. It shows that our method suppresses ringing artifacts more effectively than other mentioned methods. Furthermore, the MSE and SSIM of the proposed method in different steps for 4-time super-resolution are shown in Table 3. The result shows that each step has important contribution to the proposed method.

3.2. Study on the contribution to face recognition

Face hallucination can certainly improve the visual quality of a face image. Is the face hallucination also necessarily helpful for improving the face recognition performance when the probe and gallery images are under different resolutions? In this section we will study this issue.

In a face recognition system, when a probe image $P$ and a gallery image $G$ are under different resolutions, e.g., $G$ is with higher resolution. To do the matching between $P$ and $G$, we may employ two operations, i.e., (1) down-sampling $G$, or (2) doing face hallucination for $P$ so that $P$ and $G$ have the same resolutions. Intuitively, down-sampling $G$ will lose some useful information for recognition, and the hallucination for $P$ may keep these information and improve the recognition performance. We would
like to conduct face recognition experiment to verify this observation. It should be noted that in [24–26], some methods for improving the recognition performance on LR images have been presented, but they in fact implement an “implicit super-resolution” rather than real super-resolution reconstruction. These works will not be considered in this paper.

3.2.1. Experimental setting

The CAS-PEAL-R1 database was also used in the experiment. The data set used for face hallucination training were the same as that used in the last
experiment. For face recognition, we set up three data sets, i.e., training set, gallery set, and probe set. The training set contains 1200 images of 300 individuals for 4 images per person, which were randomly selected from the database. The gallery set contains the whole Normal subset with 1040 images of 1040 individuals. The probe set consists of 100 images of 100 individuals randomly selected from the Background and Aging subsets which have variations in image background and photographing time, respectively. All the images in the gallery and training sets were excluded from the probe set. The original HR images in probe set were down-sampled to LR images with magnification factor $S=4$ (or $S=8$), and then were respectively hallucinated by using different methods. Then, the LR images, the hallucinated images, and the original HR images were respectively used for probing. When using LR images as probe ones, the images in gallery set were also down-sampled to LR ones. Correspondingly, we have three match modes, i.e., “LR-to-LR”, “Hallucination-to-HR”, and “HR-to-HR”.

Six algorithms, namely Eigenface [27], Fisherface [28], Kernel Principal Component Analysis (KPCA) [33], Local Binary Patterns (LBP) [29], Histograms of Oriented Gradient (HOG) [30], and 2D Gabor [34] were respectively used for facial feature extraction, where the former three extract holistic features while the latter three extract local features. The nearest neighbor (NN) classifier was used for classification. The performance of face recognition was evaluated by the Cumulative Match Curve (CMC) [31].

### 3.2.2. Results

The recognition results are shown in Figs. 9–14. As shown, for each case the “HR-to-HR” mode gets significantly the best result among three modes. According to the results shown in Figs. 9–11, when using the holistic descriptor Eigenface, Fisherface, or KPCA, the “Hallucination-to-HR” mode performs equal or even worse compared to the “LR-to-LR” mode, but the situation reverses when using the local descriptor LBP or HOG as shown in Figs. 12 and 13. As a multi-scale descriptor, the 2D Gabor attains a medium result. As shown in Fig. 14, using 2D Gabor features the improvement of recognition performance is slight under 4-time super-resolution situation, but evident under 8-time super-resolution situation. Furthermore, in most cases, our method and Ma et al.’s attain the most obvious improvement for recognition performance, while Baker and Kanade’s performs worst among the compared hallucination methods.

Accordingly, we can summarize some conclusions as follows:

1. The face hallucination makes little even negative contribution to face recognition in the case of using a holistic-feature descriptor (e.g., Eigenface, Fisherface, or KPCA). The main reason is that in holistic analysis such as PCA, the low-frequency information is mainly considered, while the high-frequency components are always ignored. Therefore, low-resolution images are
enough. Furthermore, a face hallucination operation may also make some useful information lost.

(2) An efficient face hallucination is useful for improving the face recognition performance when using a local-feature descriptor (e.g., LBP or HOG). A good face hallucination algorithm can effectually recover the high-frequency information for HR image, which is quite useful for local-feature descriptor.

(3) By using the multi-scale Gabor features, the contribution of face hallucination to face recognition is also obvious, but not so effective in a smaller magnification super-resolution situation.

(4) By employing existing face hallucination algorithms, the face recognition performance of "Hallucination-to-HR" mode is still far away from that of "HR-to-HR" mode.

(5) A face hallucination algorithm that achieves better visual quality for hallucinated image does not necessarily contribute more to face recognition. For example, the Baker and Kanade's algorithm gets better image quality but worse recognition result than bilinear interpolation.

3.3. Simultaneous super-resolution and expression normalization

In this section we evaluate the extension of our method to implement super-resolution and expression normalization simultaneously. 240 images from the Expression subset were used to form the probe set for face recognition. The rest of this subset is used for training the proposed method for simultaneous super-resolution and expression normalization. Other settings were kept the same as that in the previous section. We mainly report the result of the proposed method because other face hallucination methods are not suitable for expression normalization. The image quality and the recognition performance are shown in Figs. 15 and 16. It shows that our method can effectually attain a HR image with neutral expression from a LR image with varying expression. The face recognition rate (note that all gallery images are with neutral expression) under employing our method is around 15% higher than the bilinear interpolation, and around 5% higher than using original HR images with varying expressions. We also investigated the reason of failure for classification with the LR input image. Some failed examples are shown in Fig. 17. As shown, the proposed method normalizes the expression well, however, may also distort some local facial features when the expression varies too much. The results indicate that the performance of our method is limited for the face images with large variation in facial expression.

Fig. 17. Examples of simultaneous super-resolution and expression normalization for failed recognition. The top: LR inputs with varying expression; The middle: results of simultaneous super-resolution and expression normalization; The bottom: the original HR image with neutral expression.

Fig. 18. Examples of HR hallucination with expression normalization on LR faces in realistic scenarios. In each scene, the frontal faces are located and then processed by our proposed algorithm. The processed results are displayed in the right of the scene images.
These results verify the effectiveness of the proposed method for simultaneous super-resolution and expression normalization.

3.4. Experiments on realistic scenario

We also evaluated the proposed method on the images of realistic scenarios downloaded from the Google engine. In each scene, the frontal faces were cropped manually and registered into $32 \times 24$ pixels. Then, these cropped LR images were processed with simultaneous super-resolution (into $128 \times 96$ pixels) and expression normalization by our method. Some results have been shown in Fig. 18. In realistic scenarios, face images might be various in complicated out-door lighting and adornments (e.g., eyeglasses), which makes the face hallucination far more challenging than that in the lab scenario. However, we are pleased to find that the performance of our algorithm is fairly good even in the non-lab environment. Due to the lack of corresponding HR images, the face recognition and quantitative evaluation for image quality is unavailable for these images from real scenes.

4. Conclusion

Morphological Component Analysis (MCA) is a novel tool for separating a signal into multiple semantic components in accordance with specific dictionaries. In this paper, a three-step face hallucination method has been present, where the MCA is employed to estimate a global HR face image from a LR one. Furthermore, thanks to the multi-channel decomposition capability of MCA, the proposed method can be extended to perform face hallucination and expression normalization simultaneously, where three dictionaries are designed for modeling the global face, unsharp mask, and expression, respectively. Experimental results have demonstrated the effectiveness of our method for the images from both lab environment and realistic scenarios.

Based on the experimental result, we also study the contribution of face hallucination to face recognition. The main point is that an efficient face hallucination is useful for improving the face recognition performance when using a local-feature descriptor (e.g., LBP), but makes little even negative contribution in the case of using a holistic-feature descriptor (e.g., Eigenface).

As one of the most important issues in MCA-based image decomposition, a more sophisticated dictionary design will be focused in our future work.

Acknowledgment

This work was supported by the 12th Five-year Plan China S & T Supporting Programme (No. 2012BAK16B06), NSFC (Grant No. 61172084 and 61128009), the GuangDong Program (Grant No. 20108031000004), and the Guangdong Natural Science Foundation (No. S2011040000433). Portions of the research in this paper use the CAS-PEAL face database collected under the sponsor of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

References


