Multiple-View Face Hallucination by a Novel Regression Analysis in Tensor Space

Parinya Sanguansat
Faculty of Engineering and Technology, Panyapiwat Institute of Management, Thailand

Abstract: In this paper, the novel multiple-view face hallucination method was proposed. This method is reconstructed the high-resolution face images in various poses (normal, up, down, left, and right) from a single low-resolution face image within these poses. There are two steps in our proposed method. In the first step, a high-resolution face image in the same view of the observation is reconstructed by the position-patch face hallucination framework with the improved Locally Linear Embedding (LLE), which the number of neighbours is adaptive. In the second step, the reconstructed image is used to generate the high-resolution of the other views by the novel tensor regression technique. The experimental results on the well-known dataset show that the proposed method can achieve the better quality image than the baseline methods.

Keywords: Face hallucination, tensor regression, multiple views, super-resolution.

Received January 6, 2014; accepted December 16, 2014; Published online December 23, 2015

1. Introduction

Traditionally, the interpolation-based can be used to synthesize the High Resolution (HR) image from the Low Resolution (LR) image but the results suffer from severe blurring problem. Recently, reconstruction-based [4, 6, 9, 18, 27] and learning-based [1, 3, 7, 8, 10, 11, 12, 13, 16, 20, 24, 25, 26] super-resolution algorithms are proposed to deal with this problem. The former reconstructs a HR image from multiple LR images, which are captured from the same scene with slightly different transformations. The latter generates HR image from only single LR image by learning from a set of training samples which are collected from many different subjects within same domain. These approaches, which are applied to human face domain, are also known as the face hallucination. Normally, the learning-based method can perform better but the results depend on dataset.

The serious problem of face hallucination is the reconstructed HR face image is only limited to a same view of LR face input, which is normally be a frontal view. However, the captured LR face images in real conditions are usually non-frontal view because of the camera view angle and other uncontrollable image acquisition conditions. Therefore, the solution of this problem is the important key for practical implementation. In [22] the multi-view face hallucination framework was proposed by generating LR multi-view face images from a LR single view face input and then hallucinating their corresponding HR images of these LR images.

This paper focuses on the learning-based technique for synthesizing the HR face image from a single LR image. After the HR face image of the same view as the LR one was reconstructed, the other views can be synthesized by the novel regression technique. For the first part, the position-based approach [23] is applied by the improved Locally Linear Embedding (LLE). For the second part, the novel regression analysis can be used to learn the relation between face views. The traditional regression technique is based on scalar or vector objects which the spatial information could be ignored. Classically, image must be previously vectorized before regression analysis that leads to the curse of dimensionality, small sample size, and singularity problems. To solve these problems, the novel regression analysis for tensor object is proposed in this paper to deal with the high dimensional space and multiple modalities. Combining these two parts, this framework can generate HR face images in multiple views from a single view LR face input.

This paper is organized as follows: First, the related works are discussed in section 2. Next, the Adaptive Locally Linear Embedding (ALLE) is introduced in section 3 and the novel regression technique for tensor object is proposed in section 4 followed by the proposed multi view face hallucination in section 5. Finally, the experimental results and conclusions are presented in sections 6 and 7, respectively.

2. Related Works

This paper focuses on the image enhancement in specific domain which is human face image. Face Super Resolution (SR), also known as face hallucination, aims to recover high quality, HR images of human faces from low-resolution, blurred, and degraded images or video. The previous works in face hallucination algorithm can be classified into three types. The first type is interpolation-based algorithms such as bilinear, nearest neighbour or bicubic.
interpolations. These algorithms suffer from severe blurring problem. Especially when the resolution of the input is very low, it is usually poor since no new information is added in the process. The second type is reconstruction-based method [4, 5, 6, 9, 18, 27] which try to model the process of image degradation to build the relationship between LR and HR images. This is based on reconstruction constraints and smoothness constraints. The algorithm is quite limited by the number of LR inputs which usually cannot work well in single-image super-resolution problem. The last algorithm is learning-based method [1, 3, 7, 8, 10, 11, 12, 13, 16, 20, 24, 25, 26] which becomes very popular because it can achieve higher magnification factor and better output results, especially for single-image super-resolution problem. In this method, the images of the training set contain high and LR image pairs, which are used to learn the high and LR relationship. The single image from difference subject can be enhanced by inverse of this relation.

For face hallucination, the learning-based method is preferred because the human faces have a common characteristic. Since, face images are well structured and have similar appearances; they span a small subset in the high dimensional image space. In the study by Penev and Sirovich [14] face images are shown to be well reconstructed by Principal Component Analysis (PCA) representation with 300 to 500 dimensions. This technique was widely used in face recognition tasks in the name of Eigen face [17, 19] or combined with Linear Discriminant Analysis (LDA) in fisher face [2]. Eigen transformation [24] is proposed to fit the input face image as a linear combination of the low-resolution face images in the training set. HR image was then rendered by replacing the LR training images with their HR correspondences, while retaining the same combination coefficients. Currently, the position-patch based face hallucination was proposed in [23]. This method divides a single image into multiple patches based on their positions. Only the patches in same position are used to learn HR-LR relation by LLE [15]. HR image obtained from their proposed method was generated by using the same position image patches of each training images. Therefore, the face alignment is very important in this method. The result of this work showed the improved image quality than other methods, including Eigen transformation.

However, the performance of LLE is related to the number of neighbours. The original LLE used a fixed number of neighbours for every point in space to determine the optimal weights. In this way, the weights will be not good for all points because the distribution of the samples in real data is not uniform. To solve this problem, the number of neighbours should be adaptive for each point. In this way, the ALLE is introduced by using a threshold of similarity for selecting the neighbours of each point. The result of ALLE showed that the quality of hallucinated face image is better than the original LLE.

For multiple views face hallucination, the position-patch based face hallucination has been extended to multi view face hallucination in [22] by two steps approach. The first step is to generate multiple views LR faces from a single view LR input. The second step is to obtain HR correspondences of each view. In the first step, LLE is applied to determine the reconstruction weights in LR space of the same view as input. After that, these weights are applied to generate LR face in other views. Since, there is only one view input therefore it is not possible to determine the reconstruction weights in other spaces of different views. For the second step, the position patch method is used to reconstruct HR images for each view. The drawback of this method is in the first step because the generated LR faces in all views use the same reconstruction weights only from a single view of the input. In this paper, the Tensor Regression Analysis (TRA) is proposed to solve this problem. The relationships among multiple views can be estimated by TRA. In this way, LR face images in other views will be better generated.

3. Adaptive LLE

In this section, ALLE is presented. The original idea of LLE is to discover the structure under the assumption that the relation of data is nearly linear on small areas. The objective of LLE is to minimize the reconstruction error of the set of all local neighborhoods in the input space. The error can be found by:

$$e_i = \sum_{j=1}^{N} d_{i,j} 
$$

(1)

Where $d_{i,j}$ is the $j$th neighbor of $x_i$, $k$: Is the number of the neighbors, and $N$: Is the number of training samples. This cost function is minimized under two constraints:

1. $w_{ij}$ is set to be zero if $x_j$ is not a neighbor of $x_i$.
2. $\sum_{j=1}^{k} w_{ij} = 1$.

However, the number of the neighbors $k$ is fixed to every point in space. Instead of this, the ALLE uses a threshold of similarity for selecting the neighbors of each point. The threshold of similarity $\theta$ is defined as:

$$\theta \geq \frac{\sum_{j=1}^{k} d_{i,j}}{\sum_{j=1}^{N} d_{i,j}}$$

(2)

Where $d_{i,j}$: Is the distance between input and the $j$th training sample in which $d_1 \leq d_2 \leq d_3 \leq ... \leq d_k$. In this way, the number of the neighbors $k$ of each input can be varied by this threshold.
Applying LLE or ALLE to face hallucination is easy. LR space is used to determine the weights in Equation 1 and then apply these weights to the corresponding images in HR space to reconstruct HR of the input image.

4. Tensor Regression Analysis

In this section, the proposed TRA is presented. Normally, in image processing, the relation between the set of image pairs \((a, b)\) can be determined by linear regression analysis as follows:

\[
\beta = (A^T A)^{-1} A^T B
\]  

(3)

Where \(\beta\) is the regression coefficient matrix that represents the relation between image data sets \(A=\{a_1, a_2, a_3, \ldots\}\) and \(B=\{b_1, b_2, b_3, \ldots\}\). It should be noted that, \(a\) and \(b\) are column vectors. That is the image must be previously transformed to vector by column-stack vectorization. This method is not suitable for image data because the spatial information will be lost and the dimension of vector will be very high.

In this paper, the TRA is proposed based on the least square method. That means the image can directly compute in matrix from, which is the second-order tensor, without vectorization process. Given \((X, Y)\) is the pair of the second-order tensor which has relation as:

\[
Y = X \times R^T \times L^T + E = L^T X R + E
\]  

(4)

The problem is the determination of regression coefficient matrices \(L\) and \(R\) that obtain the minimum error \(E\) by:

\[
\{L, R\} = \text{argmin}_{L, R} \sum_{n=1}^{N} \|Y_n - L^T X_n R\|^2
\]  

(5)

Where \(\|\cdot\|_F\) is the Frobenius norm which defined by \(\|X\|_F = \sqrt{tr(X^T X)} = \sqrt{tr(XX^T)}\). Therefore, the objective function should be defined as:

\[
f(L, R) = \sum_{n=1}^{N} \|Y_n - L^T X_n R\|^2
\]  

(6)

Consider:

\[
\frac{\partial f(L, R)}{\partial L} = \sum_{n=1}^{N} \frac{\partial}{\partial L} \|Y_n - L^T X_n R\|^2
\]

\[
= \sum_{n=1}^{N} \frac{\partial}{\partial L} \|Y_n - L^T X_n R\|^2
\]

\[
= \sum_{n=1}^{N} \frac{\partial}{\partial L} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2
\]

\[
= \sum_{n=1}^{N} \frac{\partial}{\partial L} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial L} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial L} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial L} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial L} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

Forcing \(\frac{\partial}{\partial L} f(L, R) = 0\), we have:

\[
\hat{R} = \left( \sum_{n=1}^{N} X_n^T LL^T X_n \right)^{-1} \sum_{n=1}^{N} X_n^T LY_n
\]  

(7)

Under the same manner, we obtain that:

\[
\frac{\partial f(L, R)}{\partial R} = \sum_{n=1}^{N} \frac{\partial}{\partial R} \|Y_n - L^T X_n R\|^2
\]

\[
= \sum_{n=1}^{N} \frac{\partial}{\partial R} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2
\]

\[
= \sum_{n=1}^{N} \frac{\partial}{\partial R} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial R} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial R} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

\[
= \sum_{n=1}^{N} \left( \frac{\partial}{\partial R} \sum_{i=1}^{N} \left( Y_{n,i} - \sum_{j=1}^{M} L_{n,j} X_{n,j} R_{n,j} \right)^2 \right)
\]

Forcing \(\frac{\partial}{\partial R} f(L, R) = 0\), we have:

\[
\hat{L} = \left( \sum_{n=1}^{N} X_n^T RR^T X_n^* + \alpha I \right)^{-1} \sum_{n=1}^{N} X_n^* RY_n^*
\]  

(9)

some applications, the solutions in Equations 8 and 10 cannot be determined because the inverse matrices are not valid. Tikhonov regularization was introduced for these ill-posed problems. In this solution, Tikhonov matrix was chosen as the identity matrix. Thus, the explicit solutions are given by:

\[
\hat{R} = \left( \sum_{n=1}^{N} X_n^T LL^T X_n + \alpha I \right)^{-1} \sum_{n=1}^{N} X_n^T LY_n
\]  

(10)

And

\[
\hat{L} = \left( \sum_{n=1}^{N} X_n^T RR^T X_n^* + \alpha I \right)^{-1} \sum_{n=1}^{N} X_n^* RY_n^*
\]  

(11)

Where \(\alpha\) is the regularized parameter.

However, Equations 11 and 12 are not the closed-form solution. The iterative method was introduced to solve this solution in Algorithm 1.

Algorithm 1: Tensor regression analysis.

Initial \(L_0\) and \(R_0\) by random values in normal distribution \(\tau \geq 0\)
Do
Compute \(R_{t+1} \leftarrow \left( \sum_{n=1}^{N} X_n^T LL^T X_n + \alpha I \right)^{-1} \sum_{n=1}^{N} X_n^T LY_n\)
Compute \(L_{t+1} \leftarrow \left( \sum_{n=1}^{N} X_n^T RR^T X_n^* + \alpha I \right)^{-1} \sum_{n=1}^{N} X_n^* RY_n^*\)
\(\tau \leftarrow \tau + 1\)
While \(\|L_{t+1} - L_t\|_2 \geq \|R_{t+1} - R_t\|_2\) and \(\|R_{t+1} - R_t\|_2 \geq \tau_R\)
Return \(L_{t+1}\) and \(R_{t+1}\)

This procedure can be extended for higher-order tensor as the same manner as High-Order Singular Value Decomposition (HOSVD) algorithm. That is taking only left projection \(L\) for the projection in each mode and repeat until it converges.

In this paper, TRA is used to determine the relation between two views. If there are \(V\) views, therefore \(V\) relations must be determined between an input view and the other views. Fortunately, the regression coefficient matrices \(L\) and \(R\) can be precomputed for the training set.

5. Multi View Face Hallucination

In this section, the proposed multiple-view face hallucination frameworks are presented. For the
baseline method in [22] there are two steps: Multi view LR faces generating step and resolution enhancement step. In the first step, suppose that \( x_v \) is an input LR face images at view \( v \), which was previously transformed into vector form. The face at arbitrary view can be approximated from linear combination of other samples at the same view as:

\[
x_v = A_{v} w_v,
\]

Where the column of matrix in \( A_{v} \) is the training LR image vector at view \( v \), and \( w_v \) is the construction coefficients which can be determined by LLE or ALLE. Based on the assumption in [22] the linear mapping between linear combination of LR faces at different views can be approximated by the same construction coefficients \( w_v \) as follows:

\[
x_u = A_{u} w_v, \quad u \neq v
\]

Where \( u \) is the other views. After obtaining LR images in all views, the position-patch based face hallucination [23] can be applied in this step to enhance the resolution for these LR images.

In this paper, the novel framework of multiview face hallucination is proposed by using TRA. From Equation 14, the linear combinations in all views are approximated by the same construction coefficients of the input view. In this way, no information from other views is used to determine these linear combinations. Instead of this, regression analysis can be used to find the relation between two views, as follows:

\[
X_u = L^*_u X_v R_{uv}, \quad u \neq v
\]

Where \( L^*_u \) and \( R_{uv} \) are the regression coefficient matrices for the relation between view \( u \) and view \( v \), which obtained from the algorithm in Table 1 and \( x_v \) is a face image matrix at view \( v \). The diagram of the proposed approach is presented in Figure 1.

6. Experimental Results

In this section, the experimental results of the proposed method are presented. All experiments were used the face images from CAS-PEAL face database [21]. There are 1,040 subjects with 21 different poses in this database but only 5 poses (up, down, left, right and frontal poses) of normal face are considered in this paper. All face images were manually aligned, which the centers of eyes and month are the control points, and cropped to 128×96 pixels for HR image and resized to 32×24 for LR image. The experiments were divided into two parts. Firstly, the performance of multiple-view face estimation was investigated. Each pose was used to generate the other poses in the same spatial resolution. Secondly, the face hallucination with multiple-view estimation was examined. The single view LR image was used as the input for reconstructing the HR image of the other views.

6.1. Multiple-View Estimation by TRA

The purpose of these experiments is to explore only the performance of the TRA for estimating the other views from a single view input, therefore this estimation was performed on the same resolution as the input without resolution improvement. The patch-based method is applied here where the patch block is 3x3 pixels with 2 pixels overlapping. For training stage, 1010 subjects are used as the training set for calculating the regression coefficient matrices. The results for high- and low-resolution are shown in Figures 2 and 3, respectively.

The results as shown below that the face area can be estimated better than the non-face area. The detail of non-face area is similar to the same area of the input because this area is various details and impossible to estimate the relationships among views.
Figure 3. Multiple-view estimation by TRA in low-resolution: The images in diagonal are the input and the rest images in each row are the estimated other views.

6.2. Multiple-View Face Hallucination by ALLE

The performance of overall process was explored in this section. The first step, a single view LR face image (32x24 pixels) is reconstructed to HR face image (128x96 pixels) of the same view as the LR input. After that, the reconstructed image is used as the input for multiple-view estimation as same process as described in subsection 6.1. For parameters setting in all experiments, the similarity threshold $\theta$ for ALLE is set to 0.1 and the number of neighbours $K$ for LLE is set to 300. The bicubic interpolation and Ma’s method [22] are the baseline in this experiment. In the proposed method, the generated LR images for the other views are not necessary as shown in Figure 1. However, the generated LR images via Ma’s method [22] are the input for the bicubic interpolation and Ma’s face hallucination. The comparisons of different view input are shown in Figures 4, 5, 6, 7, 8. It shows that, the proposed method is closely approximate to original face image than other methods. Table 1 shows the comparison of the reconstructed form the different input views, the up view reconstructed the down view, and vice versa, will encounter with some artifacts.
7. Conclusions

In this paper, the novel multi-view face hallucination is proposed by using ALLE and TRA. ALLE is the adaptive version of LLE which the number of neighbors can be adjusted for each input by the similarity threshold. TRA is the novel regression technique for tensor object in which the image can be directly computed in matrix form. The proposed framework makes the face hallucination can be applied to the real-world application because LR face image can input in any views. The single input LR image can be hallucinated to HR images in other views. The performance of the proposed method is better than the baseline methods. Moreover, the novel ALLE and TRA can be applied to other applications.

Acknowledgements

This work has been supported by Research Grant for New Scholar (MRG5480162) from Thailand Research Fund (TRF), Commission on Higher Education (CHE), and Panyapiwat Institute of Management (PIM),

Table 1. Comparison of reconstructed results

<table>
<thead>
<tr>
<th>Input View</th>
<th>Normal</th>
<th>Up</th>
<th>Down</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Up</td>
<td>Good</td>
<td>-</td>
<td>Artifact</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Down</td>
<td>Good</td>
<td>Artifact</td>
<td>-</td>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Left</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>-</td>
<td>Good</td>
</tr>
<tr>
<td>Right</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>Good</td>
<td>-</td>
</tr>
</tbody>
</table>
Thailand. The researcher would like to thank Assoc.Prof.ChomKimpan for mentoring this research.

References


Parinya Sanguansat received his BEng, MEng and PhD degrees in Electrical Engineering from the Chulalongkorn University, Thailand, in 2001, 2004 and 2007 respectively. He is an assistant professor in the Faculty of Engineering and Technology, Panyapiwat Institute of Management, Thailand. His research areas are machine learning, image processing, digital signal processing in pattern recognition including on line hand written recognition, face and automatic target recognition.