Graph Regularized Sparse Coding for Face Hallucination

1,2Kebin Huang, 1Ruimin Hu, 1Zhen Han, 1Tao Lu, 1Jun Jun Jiang and 1Feng Wang
1National Engineering Research Center on Multimedia Software, Wuhan University, Wuhan, 430072, China
2Department of Digital Media Technology, Huanggang Normal University, Huangzhou, 438000, China

Abstract: Sparse representation based face hallucination algorithms have received increasing amount of interest recently. However, most of the existing approaches fail to consider the geometrical structure of the face data, which lead to artificial effects on reconstructed High Resolution (HR) face images. In this study, a novel sparse representation based face hallucination method is proposed to reconstruct a HR face image from a Low Resolution (LR) observation. In training stage, global and local information are used to get a more expressive HR-LR dictionary pair for certain input LR patch separately. In reconstruction stage, K selection mean constrains is used to I, convex optimization, aiming to find an optimal weight for HR face image patch reconstruction. Experiments validate the proposed method.

Key words: Face hallucination, graph regularized, sparse coding

INTRODUCTION

In nowadays, intelligent video surveillance has been widely used in many areas, such as security and protection, transportation and so on. People in video surveillance are key clue for recognition. As the people usually are far from the camera, it makes the human faces in video surveillance usually have Low Resolution (LR) in spatial. The LR image lost many details and hard to be used for recognition and other postprocesses. The face images usually need to be enhanced for recognition.

Many face image SR methods, which are also called face hallucination (Baker and Kanade, 2002) are proposed to restore the LR image. They can be divided into three categories: the global face method (Wang and Tang, 2005), the local face method (Capel and Zisserman, 2001; Chang et al., 2004) and the combination method (Liu et al., 2001; Yang et al., 2008). As the local face method can fully use the partial geometry feature of human face image, such as eyes, nose, mouth and symmetry etc, it can get a better performance compared to the global method. Also, the combination method is sensitive to noise. So, this study focus on the local face method here.

Capel and Zisserman (2001) was the first to propose the local face image SR method. It divided the face image into four key regions: the eyes, nose, mouth and cheek areas. For each area, it learns a separate Principal Component Analysis (PCA) basis and reconstructs the area separately. However, the reconstructed face images in this method have visible artifacts between different regions. Inspired by manifold learning methods, Zhang et al. (2008) proposed a Locality Preserving Projections (LPP) based face image SR method. LPP algorithm is used to select the training patch samples adaptively, which can analyze the local intrinsic features on the manifold of local facial areas. PCA algorithm is used to restore the lost high-frequency components by patch-based eigen transformation with the selected training set. Zhang et al. (2008) method achieved good performance with relative small sample and there are less artifacts in adjacent area between patches. However, the methods above are Nearest Neighbors (NN) like methods (Chang et al., 2004), which are needed to assign the number of neighbors empirically.

Yang et al. (2008) introduced the idea of sparse representation to the face image SR. It decides the number of NN automatically. Based on Yang et al. (2008) pioneering works on SR, many sparse representation based face image SR algorithms are proposed. Chang et al. (2010) proposed a face sketch synthesis method using coupled over complete dictionaries and sparse representation. Jung et al. (2011) provided a position-patch based face hallucination method using convex optimization, which obtained the optimal weights for face hallucination and achieved the best results reported in literaty.

However, Jung et al. (2011)'s method, just like most of the existing sparse representation based face hallucination methods fail to consider the geometrical structure of the face data. Studies have shown that the high dimension data usually lie on a low dimension

Corresponding Author: Kebin Huang, National Engineering Research Center on Multimedia Software, Wuhan University, Wuhan, 430072, China
manifold. By preserving the geometrical structure of the data, can improve the performance of the sparse coding. (Cai et al., 2011).

Inspired by preliminary work on geometrical structure preserving, the study proposed a novel position-patch based face hallucination method using graph constrained sparse coding.

**GRAPH REGULARIZED SPARSE CODING**

**Graph construction:** Here gives the graph construction procedure briefly.

Let \( X = [x_1, x_2, ..., x_n] \) be a set of data, \( M \) is the dimension of the data, \( Q \) is the number of the data. It can construct a graph \( G = (V, W) \). \( V = Q \) is the vertices of the graph, each point is a vertex, \( W = [w_{ij}]_{Q \times Q} \) is the weighted matrix with \( w_{ij} \) being the similarity measurement between point \( x_i \) and \( x_j \). There are many choices to define the weight matrix \( W \) on the graph. Here, adopt heat kernel weighting as follows:

\[
w_{ij} = \frac{1}{c_i} e^{-\frac{(l_{ij})^2}{2h^2}}
\]

where, \( h \) is a parameter enforcing the similarity and \( c_i \) is the normalization factor. Then the weight matrix \( W \) is:

\[
W = \begin{cases} w_{ij}, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases}
\]

where, \( N_k(x_i) \) denotes the set of \( K \)-nearest neighbor of \( x_i \). The matrix \( W \) contains the geometric structure information of the data space.

**Graph regularized sparse coding:** Let \( X = [x_1, x_2, ..., x_n] \) be the data set, \( D = [d_1, d_2, ..., d_n] \) be the dictionary matrix, \( S = [s_1, s_2, ..., s_n] \) be the coefficient matrix. Sparse coding aims to find a dictionary \( D \) and a sparse coefficient matrix \( S \) whose product can best approximate the original data matrix \( X \), which can be represented as \( \min_{S,D} \|X - DS\|_2 \). The column vectors of \( D \) can be regarded as the basis vectors and each column of \( S \) is the new representation of each data point in this new space. The objective function of sparse coding can be defined as:

\[
\min_{S,D} \|X - DS\|_2^2 + \lambda \|S\|_1
\]

where, \( \lambda \) is the regularization parameter, used to balance the weight between the approximation error and the sparseness.

In order to map the weighted graph \( G \) to the sparse representations \( S \), it assumes that if two data points \( x_i \) and \( x_j \) are close in the intrinsic geometry of the data distribution, then \( s_i \) and \( s_j \), the sparse representations of these two points with respect to the new basis are also close to each other. The \( K \) nearest neighbors graph constrain can be represented as:

\[
\sum_{i \in N_k(x_j)} w_{ij} \| s_i - s_j \|_2^2 = \| S - BW \|_F^2 - \text{Tr} (BS) - \text{Tr} (LS^T)
\]

where, \( w_{ij} \) is the weight assigned to \( s_i \) and \( s_j \) represents a sparse coefficient for a data point \( x_i \). \( L = (I-W)(I-W)^T \), \( I \) is the unit matrix, \( W \) is the weighted matrix.

By incorporating the \( K \) nearest neighbors graph constrain into the original sparse coding, the optimization problem is reformulated as:

\[
\min_{S,D} \|X - DS\|_2^2 + \lambda \|S\|_1 + \text{Tr} (BS^T)
\]

Although, formula (5) is not convex in both \( D \) and \( S \), it is convex in fixing \( D \) only or fixing \( S \) only. Hence, an iterative theme is adopted to minimize formula (5) on one variable while fixing the other one.

**FACE HALLUCINATION VIA GRAPH REGULARIZED SPARSE CODING**

Here, it shows how the graph regularized sparse coding algorithm can be used to improve the performance of the face hallucination. The method mainly contains two stages: training and reconstruction. In the training stage, a series of coupled HR-LR dictionaries are gained according to the position of face image patch via graph regularized sparse coding. In the reconstruction stage, LR dictionary is used to sparse coding the input patch first, then mapping the coefficients to the HR dictionary and reconstruct the HR patches of the input LR face image. At last, by merging all reconstructed HR patches together with overlapped in adjacent patches, it can obtain the final HR face image.

Let \( I_H = [I_{h_i}^q]^q_{q=1} \) be the HR training images, \( I_L = [I_{l_i}^q]^q_{q=1} \) be the LR training images, \( Q \) is the number of training images. \( M \) and \( N \) are the dimension of each HR and LR face image by reshaping pixels to a column, \( M = sN \), \( s \) is down sampling factor. Each image in the training sets \( I_H \) and \( I_L \) is divided into a set of \( P \) small overlapping patch sets.

According to the position of face, all training patches are classified into \( P \) subsets as \( \{I_{h_i}^q\}_{q=1}^P, \{I_{l_i}^q\}_{q=1}^P, ..., \).
\[
\{I_{pk}^1\}_{q=1}^{Q} \quad \text{and} \quad \{I_{pk}^2\}_{q=1}^{Q} \quad \{I_{pk}^3\}_{q=1}^{Q} = \ldots \quad \{I_{pk}^P\}_{q=1}^{Q} = \tau, \quad p = 1, 2, \ldots, P, \quad \text{respectively. Since all faces are aligned, the contents of the subset in the same position are very similar, which can provide more detail information.}
\]

To a given position \(p\), the HR-LR subset pair  
\(I_{pk}^H = \{I_{p}^H|1 \leq q \leq Q\} \) and  
\(I_{pk}^L = \{I_{p}^L|1 \leq q \leq Q\} \), are used for training the HR dictionary \(D_{pk}^H\) and LR dictionary \(D_{pk}^L\), respectively. Then the LR dictionary can be obtained by follow formula:

\[
\min_{D_{pk}^L} \left\| D_{pk}^L S_{pk} \right\|_F^2 + \lambda \left\| S_{pk} \right\|_1 + \beta \text{Tr}(S_{pk} MS_{pk}^T) \tag{6}
\]

Then LR patch \(I_{pk}\) can be represented sparsely as \(I_{pk} = D_{pk}^L S_{pk} SL\).

In the scenario of image SR, it assumes that the LR and HR image patches share the same underlying sparse representations. Therefore, the sparse representation of a LR image patch can be applied with the HR image patch dictionary to generate a HR image patch. That is  
\(I_{p}^H = D_{p}^H S_{p} = D_{p}^H S_{p} \).

However, there is Sparse Coding Noise (SCN) \(S_{p} = S_{p} - S_{p} \). In order to make \(S_{p} \) keep in consistent with \(S_{p} \), we two measures are adopted here. One is training two dictionaries jointly for the LR and HR image patches. The other is using K selection mean constrain to reduce the SCN.

The joint couple dictionaries training can be formulated as follows:

\[
(D_{pk}^H, D_{pk}^L, S_{pk}) = \arg \min_{D_{pk}^H, D_{pk}^L, S_{pk}} \left\| D_{pk}^H S_{pk} \right\|_F^2 + \lambda \left\| S_{pk} \right\|_1 + \beta \text{Tr}(S_{pk} MS_{pk}^T) \tag{7}
\]

where, \(P = [D_{pk}^H/\sqrt{W1}, D_{pk}^L/\sqrt{W2}]\) is the concatenated HR-LR patches. \(D_{pk}^H/\sqrt{W1}, D_{pk}^L/\sqrt{W2}\) is the coupled HR-LR dictionary trained from S. W1 and W2 are the dimensions of the HR and LR patches. \(S = [s_1, s_2, \ldots, s_N] \) is the matrix collecting sparse representation vectors as columns.

The procedure of K selection constrained sparse coding can be formulated as follows:

\[
s_{pk} = \arg \min_{s_{pk}} \| I_{pk} - D_{pk}^H s_{pk} \|_2 + \lambda \left\| s_{pk} \right\|_1 + \gamma \left\| I_{pk} - s_{pk} \right\|_2 \tag{8}
\]

where, \(\gamma\) is a constant and \(l_2\)-norm is used to measure the distance between \(s_{pk}\) and \(s_{pk}\). Here take \(l_2 = 2\). As \(s_{pk}\) is unknown, the SCN cannot be directly measured. A reasonable method to estimate \(s_{pk}\) is the mean of it, denoted by \(E[s_{pk}]\). By assuming the SCN is nearly zero mean random variable vector, it could approximate \(E[s_{pk}]\) by \(E[s_{pk}]\). Then Eq. 8 can be converted into:

\[
s_{pk} = \arg \min_{s_{pk}} \| I_{pk} - D_{pk}^H s_{pk} \|_2 + \lambda \left\| s_{pk} \right\|_1 + \gamma \left\| I_{pk} - s_{pk} \right\|_2 \tag{9}
\]

Here, use the weighted average of the selected K nearest neighbors to obtain \(E[s_{pk}]\). The weighted sparse coding mean of K selected nearest is:

\[
E[s_{pk}] = \sum_{k \in K_{pk}} \omega_k s_{pk}, \tag{10}
\]

where, \(s_{pk}\) is the sparse coding of the kth selected nearest neighboring patch to patch \(p, k = 1, 2, \ldots, K\). \(\omega_k\) is the weight. \(\omega_k\) can be set to be inverse proportional to the distance between patches \(p, k\).

**EXPERIMENTAL RESULTS**

Our experiments are conducted on AR face image database (Martinez and Benavente, 1998). To evaluate the performance of the proposed method, it is compared with traditional interpolation method and sparse representation based method Jung et al. (2011) (SSR). In the experiment, it chooses a subset consisting of 50 male subjects and 50 female subjects from session 1 in AR. For each subject, 7 images with illumination and expression changes are selected. 630 face images (90 subjects) are used for training and 70 face images (10 subjects) are used for testing. The images are cropped to 120x160 pixels and aligned by five manually selected feature points. All images are degraded by blurred (with an averaging filter of size 4x4), down-sampled (by a factor of 4 times) and added Gaussian noise with standard deviation 12. The degraded LR face images are used for forming HR-LR training face image patch pairs and used for testing LR face image. Testing images are like in Fig. 1a. While using testing image as input, it firstly amplify it to 120x160 pixels.

The HR patch size in SSR and our method are set to 8x8 pixels and the overlap between neighbor patches is 32 pixels. The HR and LR dictionaries size is 700 and they are trained from the same set as our method. The sparsity regularization parameter in SSR is set to 0.1. In our experiment, \(\lambda = 10, \beta = 0.2, \gamma = 0.03, K = 5\). The parameters are set experimentally.

Some of the experimental results on benchmark test images are shown in Fig. 1. The result of bicubic interpolation method (Fig. 1b) is very smooth, but it is not clear and hard to be recognized. There are obvious ghost and artificial effects in the results of Jung’s method (Fig. 1c). Compared with the reference methods, the results of the proposed method (Fig. 1d) remained more details information and less ghost and artificial effect.
The average values of PSNR, RMSE and SSIM (structural similarity) are used to evaluate the objective quality of the reconstructed images. The average values calculated from different methods are shown in Table 1.

It can be seen from Table 1 that the proposed method obtained the lowest RMSE (the lower the better), the highest PSNR and SSIM values (both are the higher the better). It demonstrates that the proposed method can reconstruct an image closest to the original HR image. The results between subjective quality and objective quality are consistent. It validates the effectiveness and advancement of the proposed method.

CONCLUSION

In this study, a novel sparse representation based face image SR method is proposed. It incorporates the position information of patches and the intrinsic geometric structure information between patches into the dictionaries, which makes the learned dictionary more expressive. By using K selection mean and l1 regularized least squares optimization through the learned dictionary, an optimal solution can be obtained for reconstructing. Experiments conducted on AR face image database validate the proposed method both in subjective and objective quality.

ACKNOWLEDGMENTS

The research was supported by the major national science and technology special projects (2010ZX03004001-03), the National Basic Research Program of China (973 Program) (2009CB320906), the National Natural Science Foundation of China (60970160, 61070080, 61003184, 61172173, 61231015), the Fundamental Research Funds for the Central Universities.

REFERENCES


