http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet

Asian Network for Scientific Information

308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (11): 1883-1887, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.1883.1887

Graph Regularized Sparse Coding for Face Hallucination

[1,2]Kebin Huang, [1]Ruimin Hu, [1]Zhen Han, [1]Tao Lu, [1]Jun Jun Jiang and [2]Feng Wang
[1]National Engineering Research Center on Multimedia Software, Wuhan University, Wuhan, 430072, China
[2]Department 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 l_1 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 recognization. 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 recognization.

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 literacy.

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 $X = [x_1, x_2, ..., x_N] \in R^{M\times Q}$ 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 vertice. $W = [w_{i,j}]_{Q\times Q}$ is the weighted matrix with $w_{i,j}$ 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:

$$w_{i,j} = 1/c_i \cdot e^{-\|x_i - x_j\|^2/h} \tag{1}$$

where, h is a parameter enforcing the similarity and c_1 is the normalization factor. Then the weight matrix W is:

$$W = \begin{cases} w_{i,j}, \text{ if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

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 $X = [x_1, x_2, ..., x_N] \in R^{M\times Q}$ be the data set, $D = [d_1, d_2, ..., d_L] \in R^{M\times L}$ be the dictionary matrix, $S = [s_1, s_2, ..., s_N] \in R^{L\times Q}$ 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 $\min_{D,S} \|X-DS\|$. 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:

$$\min_{D,S} \|X - DS\|^2 + \lambda \|S\|_1 \tag{3}$$

where, λ 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:

$$\sum_{i=1}^{N} \left\| s_i - \sum_j w_{ij} s_j \right\|^2 = \|S - SW\|^2 = Tr(SLS^T) \tag{4}$$

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:

$$\min_{D,S} \|X - DS\|^2 + \lambda \|S\|_1 + \beta Tr(SMS^T) \tag{5}$$

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 HR dictionary and reconstruct the HR patch 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 $I_H = \{I_{H,q}\}_{q=1}^{Q} = \{I_{H,1}, I_{H,2}, ..., I_{H,Q}\} \in R^{M\times Q}$ and $I_L = \{I_{L,q}\}_{q=1}^{Q} = \{I_{L,1}, I_{L,2}, ..., I_{L,Q}\} \in R^{N\times Q}$ represent the HR and LR training face 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, $M = s^2N$, 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 $\{I^1_{H,q}\}_{q=1}^{Q}$, $\{I^2_{H,q}\}_{q=1}^{Q}$, ...,

$\{I^P_{H,q}\}^Q_{q=1}$ and $\{I^1_{L,q}\}^Q_{q=1}$, $\{I^2_{L,q}\}^Q_{q=1}$, ..., $\{I^P_{L,q}\}^Q_{q=1}$, p = 1, 2, .., P, 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 $I^p_H = \{I^p_{H,q}|1\le q\le Q\}$ and $I^p_L\{I^p_{L,q}|1\le q\le Q\}$, are used for training the HR dictionary D^p_H and LR dictionary D^p_L, respectively. Then the LR dictionary can be obtained by follow formula:

$$\min_{D^p_L,S_L}\left\|I^p_L - D^p_L S_L\right\|^2 + \lambda\left\|S_L\right\|_1 + \beta Tr(S_L M S_L{}^T) \tag{6}$$

Then LR patch I^p_L can be represented sparsely as $I^p_L = D^p_L SL$.

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 $IpH = D^p_H S_H \approx D^p_H S_L$.

However, there is Sparse Coding Noise (SCN) $\sigma_s = S_H - S_L$. In order to make S_H keep in consistent with S_L, 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:

$$\begin{aligned}\{D^p_H, D^p_L, S\} = \arg\min_{D^p_H, D^p_L, S}\left\|I^p - D^p S\right\|^2_2 \\ + \lambda\left\|S\right\|_1 + \beta Tr(SMS^T)\end{aligned} \tag{7}$$

where, $I^p = [I^p_H/\sqrt{W1}, I^p_L/\sqrt{W2}]^T$ is the concatenated HR-LR patches. $D^p = [D^p_H/\sqrt{W1}, D^p_L/\sqrt{W2}]^T$ is the coupled HR-LR dictionary trained from S. W1 and W2 are the dimensions of the HR and LR patches. $S = [s_1, s_2, ..., s_i]$ is the matrix collecting sparse representation vectors as columns.

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

$$\begin{aligned}s^p_L = \arg\min\left\|I^p_L - D^p_L s_L\right\|_2 + \lambda\left\|s_L\right\|_1 \\ + \gamma\left\|s_L - s_H\right\|_{l_p}\end{aligned} \tag{8}$$

where, γ is a constant and l_p-norm is used to measure the distance between s_L and s_H. Here take $l_p = 2$. As s_H is unknown, the SCN cannot be directly measured. A reasonable method to estimate s_H, is the mean of it, denoted by $E[s^H_{N(k)}]$. By assuming the SCN is nearly zero mean random variable vector, it could approximate $E[s^H_{N(p)}]$ by $E[s^L_{N(k)}]$. Then Eq. 8 can be converted into:

$$\begin{aligned}s^p_L = \arg\min\left\|I^p_L - D^p_L s^p_L\right\|_2 + \lambda\left\|s^p_L\right\|_1 \\ + \gamma\left\|s^p_L - E[s^L_{N(p)}]\right\|_{l_p}\end{aligned} \tag{9}$$

Here, use the weighted average of the selected K nearest neighbors to obtain $E[s^L_{N(p)}]$. The weighted sparse coding mean of K selected nearest is:

$$E[s^L_{N(p)}] = \sum_{k\in N(p)} \omega_{p,k} s_{p,k} \tag{10}$$

where, $s_{p,k}$ is the sparse coding of the kth selected nearest neighboring patch to patch p, k = 1,2..., K. $\omega_{p,k}$ is the weight. $\omega_{p,k}$ can be set to be inverse proportional to the distance between patches p and 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 120×160 pixels and aligned by five manually selected feature points. All images are degraded by blurred (with an averaging filter of size 4×4), 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 120×160 pixels.

The HR patch size in SSR and our method are set to 8×8 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 of 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 detail information and less ghost and artificial effect.

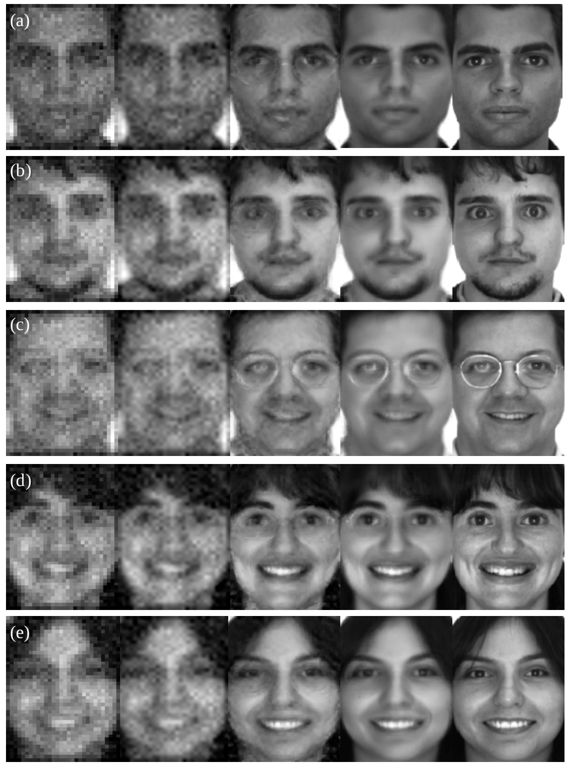

Fig. 1(a-e): Visual effects comparison between the reconstructed face images via different methods. (a) LR input face images, (b)-(d) Results reconstructed by bicubic interpolation method, SSR and the proposed method and (e) Original HR face images

Table 1: The average values of RSME, PSNR (dB) and SSIM calculated from the reconstructed images with compared methods

Method	Bicubic	SSR	Proposed method
RMSE	23.2576	10.4233	10.9425
PSNR	20.9062	27.8270	28.2910
SSIM	0.5731	0.7987	0.8605

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 showed 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 l_1 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 (2010ZX03004-001-03), 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

Baker, S. and T. Kanade, 2002. Limits on super-resolution and how to break them. IEEE Trans. Pattern Anal. Mach. Intell., 24: 1167-1183.

Cai, D., X. He, J. Han and T.S. Huang, 2011. Graph regularized non-negative matrix factorization for data representation. IEEE Trans. Pattern Anal. Mach. Intell., 33: 1548-1560.

Capel, D. and A. Zisserman, 2001. Super-resolution from multiple views using learnt image models. Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, December 8-14, 2001, Kauai, Hawaii, pp: 627-634.

Chang, H., D. Yeung and Y. Xiong, 2004. Super-resolution through neighbor embedding. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 27-July 2, 2004, Washington, DC., USA., pp: 275-282.

Chang, L., M. Zhou, Y. Han and X. Deng, 2010. Face sketch synthesis VIA sparse representation. Proceedings of the 20th International Conference on Pattern Recognition, August 23-26, 2010, Istanbul, pp: 2146-2149.

Jung, C., L. Jiao, B. Liu and M. Gong, 2011. Position-patch based face hallucination using convex optimization. IEEE Signal Process. Lett., 18: 367-370.

Liu, C., H.Y. Shum and C.S. Zhang, 2001. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, December 8-14, 2001, Washington, DC., pp: 192-198.

Martinez, A.M. and R. Benavente, 1998. The AR face database. CVC Technical Report No. 24, June 1998.

Wang, X. and X. Tang, 2005. Hallucinating face by eigentransform. IEEE Trans. Syst. Man Cybern. C, 34: 425-434.

Yang, J.C., H. Tang, Y. Ma and T. Huang, 2008. Face hallucination VIA sparse coding. Proceedings of the 15th IEEE International Conference on Image Processing, October 12-15, 2008, San Diego, CA., pp: 1264-1267.

Zhang, X., S. Peng and J. Jiang, 2008. An adaptive learning method for face hallucination using locality preserving projections. Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition, September 17-19, 2008, Amsterdam, pp: 1-8.