FACE HALLUCINATION VIA LOCALITY-CONSTRAINED LOW-RANK REPRESENTATION

Tao Lu[1,2], *Zixiang Xiong*[2], *Yongjing Wan*[1], *Wei Yang*[1],

1.School of Computer science and engineering,Wuhan Institute of Technology, Wuhan, China ,430073
2.Department of Electrical and Computer Engineering Texas A&M University College Station, TX 77843

ABSTRACT

Face hallucination (FH) based on sparse representation (S-R) and locality-constrained representation (LCR) gives reasonably good performance. However, neither SR- nor LCR-based methods make full use of the structure information in the training data. On the other hand, low-rank representation (LRR) has been utilized to cluster samples into their respective classes by exploiting low-rank structures of the data. In this paper, we propose a locality-constrained low-rank representation (LCLRR) method to take advantage of both LCR and LRR for FH. LCLRR first enforces a low-rank constraint on choosing the dictionary atoms that belong to a subspace that correspond to the same cluster, it then imposes a locality constraint on selecting atoms that are in the vicinity of test samples. Experiments show that LCLRR outperforms both SR- and LCR-based methods on subjectively and objectively, proving that exploiting the structure information in the training data is feasible in face hallucination.

***Index Terms*—** Face Hallucination, Low-rank Representation, Locality-constrained Representation,Alternating Direction Method of Multiplier

1. INTRODUCTION

Face hallucination (FH), which is also called face super-resolution, has been widely used in many applications. Inspired by the pioneering work of Baker [1], various learning-based face hallucination methods have been proposed to infer the missing information by the help of training data. Wang [2] *et al.* utilized principal component analysis (PCA) to explore *a prior* information from training samples by transferring LR features into HR ones. Chang [3] *et al.* first introduced manifold learning into super-resolution algorithm by locally linear embedding (LLE). Manifold embedding assumes that each input image can be represented by a weighted sum of certain dictionary atoms. This simple assumption leads to easy computation and often yields good results in super-resolution scenarios.

Ma *et al.* [4] proposed a position-patch based FH using all patches from the same position in a dictionary, and used a least squares representation (LSR) to obtain the reconstruction weights. Furthermore, in order to solve the over-fitting problem, sparse representation (SR) constraints [5] are embedded into manifold pursuit by adaptively selecting dictionary atoms. However, sparse constraints favor sparsity more than locality in the feature space, which often leads to unsmooth solutions. Recently, the idea of locality-constrained representation (LCR) [6] has been proposed to give more promising results for manifold based FH. The local manifold distance is used to determine weights on the representation coefficients by following the observation that nearer neighborhoods make greater contributions to the final reconstruction. Wang *at el.* [7] used l_2- or l_1-norm adaptive regularization on the representation coefficients.

Although the above approaches consider linear subspace, they ignore an important fact that patches form different classes may lie in independent linear subspaces. Structured constraints on these independent subspace should have low rank. Recently, low-rank representation (LRR) has been used for unsupervised subspace clustering [8, 9]. In [10, 11], low-rank representation was employed to exploit the structure of data by clustering the test signal into the most suitable independent subspace, achieving promising result in recognition tasks. We make the assumption that patches from different clusters belong to independent linear subspaces of low rank. The best way to utilize the structure information is to enforce the dictionary atoms from the same class with more discriminative ability.

In this paper, we propose a novel approach called locality-constraint low-rank representation (LCLRR) to cluster the input to its potential low-rank subspace and represent it based on LCR with samples from the same subspace. LRR assumes that different subspaces correspond to different patch clusters, consequently, the input patches are reconstructed by atoms belonging to just one cluster instead of a mixture of clusters. This leads to a boost in reconstruction performance in super-resolution, as shown in Fig. 1. Furthermore the reason why LCLRR works better than LCR and SR is because it has a different dictionary atom selection mechanism. Whereas L-

This work is supported by the grant of China Scholarship Council, the National Natural Science Foundation of China(61172173,61502354,61501413,U1404618), the Natural Science Foundation of Hubei Province of China (2012FFA099,2012FFA134, 2013CF125,2014CFA130,2015CFB451), the Key Project of the Educational Department of Hubei Province of China (D20141505),Scientific Research Foundation of Wuhan institute of technology.

CR and SR may choose samples from different classes (SR favors sparsity and do not care about the cluster information, and LCR emphasize the closest atoms in the subspace), L-CLRR considers both locality manifold structure and cluster constraints, with atoms from the same subspace and the nearest to the input contributing more to the reconstruction. Our experimental results confirm that low-rank constraints indeed boost the super-resolution performance.

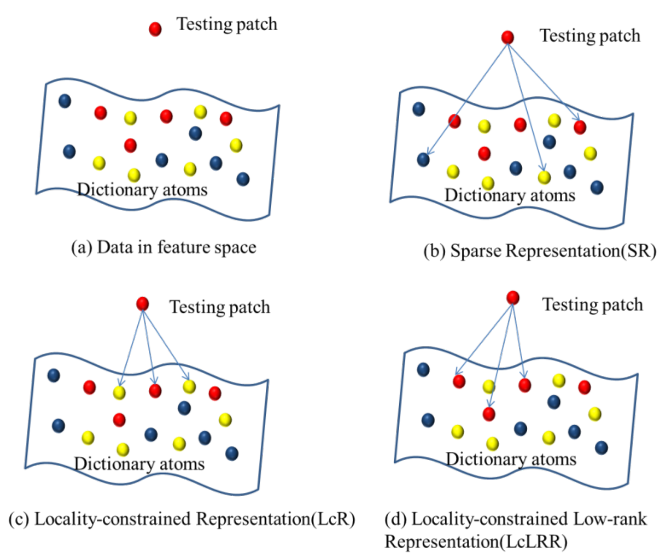

Fig. 1. Illustration of why LCLRR works superior to SR and LCR. Different colors denote atoms from different clusters. LCLRR chooses dictionary atoms from same clusters that lie closest to input testing sample.

2. RELATED WORKS

2.1. Position based FH by SR

Defining H^n as the HR training face images, $n = 1, 2, \ldots, N$. N is the number of training samples.for each image, we divided it into small overlapped patch set $\{H^n(i,j)|1 \leq i \leq U, 1 \leq j \leq V\}$. U and V are patch number in column and row respectively, term (i,j) indicates the position location. Corresponding, we down-sample and blur the HR dataset to form the LR training dataset $\{L^n(i,j)\}$. For each position (i,j) we have input patch and LR HR dictionaries as $y(i,j) \in R^{d\times1}$,$H(i,j) \in \Re^{t^2d\times N}$,$L(i,j) \in R^{d\times N}$. d is the square of patch size, t is amplification factor. SR based FH uses patches from all training samples as dictionary in same position to represent input patch $y(i,j)$. The embedding weights will be calculated by follows:

$$\arg\min_{\alpha(i,j)} \|y\,(i,j) - L\,(i,j)\,\alpha(i,j)\|_2 + \lambda\|\alpha\,(i,j)\|_0 \quad (1)$$

Here,$\|\bullet\|_0$is l_0-norm and can be replaced by l_1-norm in optimization. $\lambda$is balance parameter, $\alpha(i,j) \in R^{N\times1}$ is a vector denotes the LR representation weights for certain position input patch. Given input LR patch $y(i,j)$, the purpose of FH is to infer HR patch $x(i,j)$ then integrate all position patches to form the HR output X. Based on manifold consistency assumption, we calculate HR patch by $x(i,j) = H(i,j)\alpha(i,j)$, thus integrate all position patches to output HR X.

2.2. Locality-constrained Representation

Different from SR, LCR introduces a manifold regularization on optimization. Because all process is on image patch, so we can omit position index (i,j). The locality-constrained objectives function as follows:

$$\arg\min_{\alpha} \|y - L\alpha\|_2 + \lambda\,\|d_i \otimes \alpha\|_2^2 \quad s.t. 1^T\alpha = 1 \quad (2)$$

Here, $\otimes$ denotes element wise product, d_i is the locality adaptor that gives different freedom for each basis vector punished by the similarity to the input patch . Specially,$d_i = exp(\frac{\|y-l_i\|_2^2}{\sigma})$, l_i indicates i-th atom in LR dictionary L. σ is used to adjust the weight decay speed for the locality adaptor. Fortunately, formula (2) has analytical solution.

2.3. Low-rank Representation

Low-rank representation is widely used in data segmentation and classification tasks [8, 9]. Generally speaking,low-rank property of matrix should be thought as intrinsic structure information without noise. Consider a matrix $X = [x_1, x_2, \ldots, x_n] \in R^{M\times N}$, the columns indicate feature vectors. The objective function as follows:

$$\arg\min_{Z} \|Z\|_* \quad s.t.\ X = DZ \quad (3)$$

Where, D is the dictionary, $\|\bullet\|_*$ denotes nuclear norm which was used to relax the low-rank optimization to convex problem. Each Z_i provides a representation of the column vector x_i using the columns of the dictionary D. As we wish, the low-rank constraint on Z will automatically seek the most intrinsic subspace for each image.

3. FACE HALLUCINATION VIA LCLRR

As above analysis, LRR do benefit to reveal the intrinsic subspace about data structure. Inspired by recent works in face recognition [10, 11], we introduce low-rank constraint into manifold learning super-resolution algorithm.

3.1. Algorithm

By low-rank structure, we cluster dictionary atoms with input patch in the same subspace. In this subspace, the original input patches will be linear represented by the neighborhood from the same cluster. The ideal is to find a weighted vector(that indicates a low reconstruction error) elements of which when multiplied with the corresponding column of the dictionary L results in a low rank matrix $Ldiag(\alpha)$,and this matrix can be called as Low-rank Coding. In order to get the embedding weights, we minimize the following objective function:

$$\arg\min_{\alpha} \|y - L\alpha\|_2^2 + \lambda \|Ldiag\,(\alpha)\|_* + \beta\,\|d_i \otimes \alpha\|_2^2 \quad (4)$$

Here, α is the low-rank constrained coefficient, λ and β are balance parameters for controlling the contribution from low-rank and locality constraints, $\otimes$ denotes element wise product. $d_i = exp(\frac{\|y-l_i\|_2^2}{\sigma})$ is the locality adaptor used for measure the distance between the input patch and each dictionary atoms. For robustness, we normalize the values of d_1 between 0 to1. Nuclear norm in (3) is always used as a convex substitute for rank operation. Matrix $Ldiag(\alpha)$ represents the vectors are used to reconstruct the input patch. Minimizing the rank of this matrix means that the vector selected for reconstruct signal y using only those training patches belong to a low-rank subspace.

3.2. Optimization

Objective function (4) will be solved by alternating direction method of multiplier(ADMM) by introducing a variable $\Lambda = Ldiag(\alpha)$. The formula (4) can be written as:

$$\arg\min_{\alpha,\Lambda} ||y - L\alpha||_2^2 + \lambda||\Lambda||_* + \beta||d_i \otimes \alpha||_2^2 \;\; s.t. \Lambda = Ldiag(\alpha) \quad (5)$$

The above problem can be solved by following augmented Lagrange function.

$$\begin{aligned}&\arg\min_{\alpha,\Lambda} ||y - L\alpha||_2^2 + \lambda||\Lambda||_* + \beta||d_i \otimes \alpha||_2^2 \\ &+ tr(\Delta^T(Ldiag(\alpha) - \Lambda)) + \tfrac{\mu}{2}||Ldiag(\alpha) - \Lambda||_F^2\end{aligned} \quad (6)$$

Where tr(.) is trace operation, $\|\bullet\|_F$ is the Frobenius norm, Δ is Lagrange multiplier and μ is the penalty parameter, λ and β are parameters for balance different regularization terms. The steps of optimization is described in Algorithm1, the first step is to calculate the low-rank norm by fixed other variables. Low-rank matrix Λ will be update by follows:

$$\arg\min_{\Lambda} \frac{\lambda}{\mu}||\Lambda||_* + \frac{1}{2}||\Lambda - (Ldiag(\alpha) + \Delta/\mu)||_F^2 \quad (7)$$

This formulate can be solved by the singular value thresholding operator as [9]. And the closed form of α in step2 can be obtained by simple algebraic method.

Algorithm 1 ADMM algorithm for LCLRR

Require: $y\ D\ \lambda\ \beta$
Ensure: α
initilization $\alpha = 0, \Delta = 0, \Lambda = 0, \rho = 0, \mu = 10^6, \varepsilon = 10^{-8}$
While not converged do
1: Fix others and update Λ by formulate(7)
2: Fix others and update α by
$\alpha = (P + diag(p_1)/p_2)$
$P = 2L^TL + 2\beta diag(d) \otimes diag(d)$
$P_1 = \mu(\Lambda \otimes L)^T 1$
$p_2 = 2L^Ty + \mu(\Lambda \otimes L)^T 1 - (\Delta \otimes L)^T 1$
3: Update Lagrange and penalty parameters
$\Delta \leftarrow \Delta + \mu(Ldiag(\alpha) - \Lambda)$
$\mu \leftarrow min(\rho\mu, \mu)$
4: Check for convergence
$||Ldiag(\alpha) - \Delta||_\infty < \varepsilon$
end while

4. EXPERIMENTAL RESULTS

4.1. Database and Parameter Setting

We conducte simulation experiments on FEI Database [12]. This database contains 400 images which are cropped to 120×100 pixels, and 360 images are randomly selected to make up the training set, the rest 40 images are used for testing. In order to evaluate the effectiveness of our algorithm, we compare our results with some state-of-the-art algorithms such as LLE [3], LSR [4], SR [5], LCR [6]. LR images are smoothed by an average filter and down-sampled by a factor of 4, than the size of LR images is 30×25 pixels according to HR images. For fairly comparison, we set the parameters for their best performance from their papers for the other methods. We set HR patch size to 12×12 pixels with overlapped 4 pixels, and corresponding LR patch size as 3×3 pixels with overlapped 1 pixel. For Chang 's neighbor embedding method, the number of neighbors is set to 100. For Jung 's SR method, error tolerance is set to 1.0 during seeking sparse optimal solution. For Jiang 's LCR method, balance parameter is set to 0.04 as in the paper for best performance. In our method, we set HR and LR resolution and overlapped pixels same as above patch-based methods, thus we fine tune the low-rank and locality balance parameters $\lambda = 0.1$ and β=0.004.

4.2. Experimental results

As usual, we adopt the same assessment method such as Peak signal-to-noise ratio (PSNR) and Structural similarity (SSIM) to measure the objective and subjective quality of the reconstructed images with other algorithms. As shown in Fig.2, we list four facial images for comparing the subjective quality of each algorithm. From the Fig.2, we can find that our method is better than other methods in subjective quality (enlarger image for easy observation). LLE method suffered by under-fitting, the reconstructed facial image has blur effects in details of mouth and eyes. LSR method thus suffered by over-fitting and also not enough detail about facial location.

Fig. 2. Comparison of results based on different method.(a)Input LR faces. (b)LLE[3]. (c)LSR[4]. (d)SR[5]. (e)LCR[6].(f)our method. (g)Original HR faces.

SR and LCR methods have similar subjective performance, thus our method yields more detail in mouth and eyes regions, and are more identical with original images. Furthermore, in Fig.3, PSNR and SSIM of all 40 testing images are listed. we can see that our method has better PSNR and SSIM values than other methods. The average PSNR(SSIM) of our method is 1.26 db(0.0248) higher than that of the LLE method, 1.12 db(0.0159) higher than the LSR method, 0.91 db(0.0142) higher than that of the SR method, and 0.38db(0.0078) higher than that of the LCR method. Therefore, we can conclude that our method has better performance both in objective and subjective rather than some state-of-the-art algorithms.

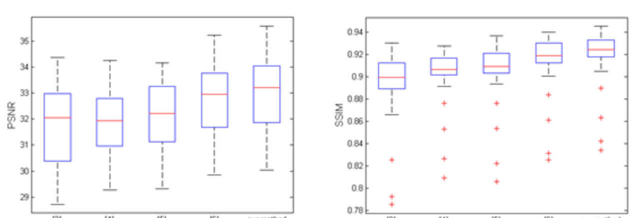

Fig. 3. Comparisons of PSNR and SSIM for different methods.The average of PSNR and SSIM of different methods: LLE [3](PSNR=31.75, SSIM=0.8945), LSR [4](PSNR=31.90, SSIM=0.9034), SR [5](PSNR=32.11, SSIM=0.9051), LCR [6](PSNR=32.63, SSIM=0.9115), the proposed LCLRR method(PSNR=33.02, SSIM=0.9193).

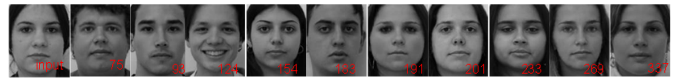

Fig. 4. A cluster contains top 10 similarity images are listed. Red number is index of samples.

4.3. The role of Low-rank representation

In order to test the role of LRR in the proposed algorithm, we take a simple test on how different algorithms to choose suitable dictionary atoms to reconstruct the input image. We set the first test facial image as input image, then we use k-means clustering algorithm to cluster all the training samples to form a cluster, because we use facial image as dictionary atoms directly, so the representation coefficients can index which image will be selected. As shown in Fig.4, 10 facial images(with most similarity to input image) which assumed into a cluster, the red number is index of the samples. As shown in Fig.5.

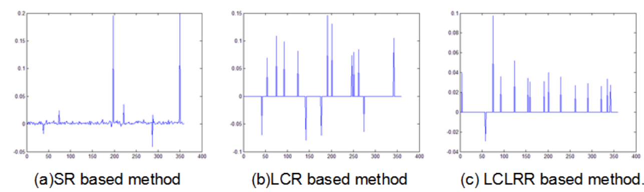

Fig. 5. The weights vector of input image. The nonzero elements indicate the selected atoms

We list the weights vector for three different algorithms, in order to observe, we cut down many small weights for focusing which atoms will be chosen, we can find that the SR method can select 6 neighbors from the cluster, the LCR can select 7 neighbors from the cluster, and our method can select 9 neighbors from the cluster. Therefore, LRR plays reasonable role in clustering dictionary atoms while coding, thus LCR method can also be treated as a special case of our method.

5. CONCLUSION

In this paper, we propose a low-rank representation based face hallucination algorithm. Rather than only focus on the locality manifold regularization on representation, we take both advantage of low-rank and locality constraints on representation, enforcing to choose the dictionary atoms from same cluster with input signal and considering the locality manifold distance. Experimental results shown our method have both better subjective and objective quality than some state-of-the art algorithms. Nonetheless, in future, we will focus on how to improve the robustness for input noise with low-rank representation method.

6. REFERENCES

[1] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 9, pp. 1167–1183, Sep 2002.

[2] Xiaogang Wang Xiaogang Wang and Xiaoou Tang Xiaoou Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, 2005.

[3] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 1, no. 3, pp. I–I, 2004.

[4] Xiang Ma, Junping Zhang, and Chun Qi, "Hallucinating face by position-patch," *Pattern Recognition*, vol. 43, no. 6, pp. 2224–2236, 2010.

[5] Cheolkon Jung, Licheng Jiao, Bing Liu, and Maoguo Gong, "Position-patch based face hallucination using convex optimization," *Signal Processing Letters, IEEE*, vol. 18, no. 6, pp. 367–370, June 2011.

[6] Junjun Jiang, Ruimin Hu, Zhongyuan Wang, and Zhen Han, "Noise Robust Face Hallucination via Locality-Constrained Representation," *Multimedia, IEEE Transactions on*, vol. 16, no. 5, pp. 1268–1281, 2014.

[7] Zhongyuan Wang, Ruimin Hu, Shizheng Wang, and Junjun Jiang, "Face Hallucination Via Weighted Adaptive Sparse Regularization," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 5, pp. 802–813, 2014.

[8] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust Recovery of Subspace Structures by Low-Rank Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 171–184, 2013.

[9] Kewei Tang, Risheng Liu, Zhixun Su, and Jie Zhang, "Structure-Constrained Low-Rank Representation," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 25, no. 12, pp. 2167–2179, 2014.

[10] D Arpit, G Srivastava, and Yun Fu, "Locality-constrained Low Rank Coding for face recognition," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 1687–1690.

[11] Lihong Peng Ziheng Jiang, Ping Guo, "Locality-Constrained Low-Rank Coding for Image Classification," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014, pp. 2780–2786.

[12] Carlos Eduardo Thomaz and Gilson Antonio Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, 2010.