FACE HALLUCINATION VIA CAUCHY REGULARIZED SPARSE REPRESENTATION

Shenming Qu[1,2,3], *Ruimin Hu*[1], *Shihong Chen*[1], *Zhongyuan Wang*[1], *Junjun Jiang*[1], *and Cheng Yang*[1]

[1] National Engineering Research Center for Multimedia Software,
Computer School of Wuhan Univ., China
[2] Research Institute of Wuhan University in Shenzhen, China
[3] School of Software, Henan Univ., China

ABSTRACT

In dictionary-learning-based face hallucination, the testing image is represented as a linear combination of the training samples, and how to obtain the optimal coefficients is the primary issue. Sparse representation (SR) has ever been widely used in face hallucination, however, due to the fact that SR overemphasizes the sparsity, the obtained linear combination coefficients turn out far aggressively sparse, then leading to unsatisfactory hallucinated results. In this paper, we present a moderately sparse prior model for face hallucination problem with the $L1$ norm penalty in classic SR replaced by a Cauchy penalty term. An iterative optimization is further presented to solve the minimization of Cauchy regularized objective function. The experimental results on public face database demonstrate that our method is much more effective than state-of-the-art methods.

***Index Terms*—** Super-resolution, face hallucination, sparse representation, Cauchy regularization

1. INTRODUCTION

Face hallucination, or face super-resolution, has recently become a hot topic in video applications, such as video surveillance, due to the increasing number of practical applications of the algorithms proposed [1–10]. It is universally acknowledged that face hallucination is cast as an inverse problem, it recovers the original high-resolution (HR) image from the low-resolution (LR) input, and this can be represented by the observation model as follows:

$$y = Hx, \tag{1}$$

where y is the observed LR image (column-stacked), x is the unknown HR image (column-stacked) to be estimated, $y \in R^N$, $x \in R^M$ and $N < M$. The matrix $H \in R^{N\times M}$ represents the imaging system, consisting of several processes, such as warping, blurring, down-sampling and more. Given one LR observation y, face hallucination is to solve the above-mentioned inverse problem to obtain an approximation $\hat{x}$ to the unknown HR image x. Since the $N \times M$ matrix H has far fewer rows than columns, the inverse problem is underdetermined, infinitely many HR images satisfy the above reconstruction constraint. Thus, to recover a visually pleasing HR image, various regularization approaches have been proposed to employ some image prior to stabilize the inversion of this ill-posed problem [11–14].

Recently, as a powerful tool for statistical signal modeling, SR has been used as a way of forming regularization in inverse problems. Candes et al. [15] used an iterative procedure to get more sparsity solution for sparse signal recovery. Lately, Yang et al. [16] are the first to introduce $L1$-norm SR to face hallucination, who enforced corresponding LR and HR patches to share the same SR to enhance the detailed facial information. Jung et al. [8] advanced a position-patch face hallucination method to the sparsity constraint least square problem, and showed state-of-the-art performance in face hallucination. Very recently, Dong et al. [17] proposed a method to explore the image nonlocal self-similarity, they utilized non-locally centralized sparse representation (NCSR) to obtain good estimates of the sparse coding coefficients.

However, due to SR based methods are in favor of sparsity, the obtained linear combination coefficients turn out far aggressively sparse, those methods may select very distinct basis images to reconstruct the input image, which will result in unsatisfactory hallucinated result. In fact, SR based method is to take a sparse constrained optimization to replace the least square estimation, and to obtain more suitable solution. Whereas, Laplacian prior assumed for $L1$-norm may not quite agree with the actual distribution.

In this paper, we manage to seek a more fitted prior model for hallucinating face images, we propose a so-called Cauchy Regularized Sparse Representation (CSR) model for face hallucination to improve the effectiveness of SR. By placing a Cauchy prior on solution, we can derive a moderately s-

The research was supported by the National Natural Science Foundation of China (61172173, 61303114,U1404618), Guangdong-Hongkong Key Domain Breakthrough Project of China (No. 2012A090200007), the National High Technology Research and Development Program of China (863 Program 2013AA014602),the Fundamental Research Funds for the Central Universities (2042014kf0286), Science and Technology Foundation of Guizhou Province (No. LKS[2011]1).

parse regularization method, and our method can surpasses $L1$-norm SR based methods.

In contrast to Ref. [2, 4–8], the main contribution of this paper is as follows:
(1) We are the first to introduce Cauchy distribution to model coefficient prior for face hallucination, and achieve satisfactory hallucinated results.
(2) We use an iterative procedure to gradually approaching the optimal linear combination coefficients.

The rest of this paper is organized as follows. Section 2 presents the proposed CSR model. Section 3 conducts the experiments on FEI face database [18], and Section 4 concludes the paper.

2. PROPOSED METHOD

For face hallucination, the traditional sparse coding model is equivalent to the following optimization problem:

$$J(\alpha) = ||y - D\alpha||_2^2 + \lambda \sum\nolimits_i |\alpha_i|, \tag{2}$$

where $y = [y_1; y_2; ...; y_N] \in R^N$ is the input LR face image to be coded, $D = [d_1; d_2; ...; d_M] \in R^{N \times M}$ is the training dictionary with column vector d_j being the jth basis images, and the solution denoted by M-dimensional vector α consists of a set of linear combination coefficients. Each entry in α is associated with an individual base in the training dictionary.

2.1. Cauchy Regularized Sparse Representation (CSR)

To estimate those linear combination coefficients, the classic $L1$-norm SR assumes that coefficient vector α obeys i.i.d. zero-mean multivariate Laplacian distribution, namely,

$$P_L(\alpha) = \frac{1}{(2\mu)^M} \exp(-\frac{||\alpha||_1}{\mu}), \tag{3}$$

where scale parameter $\mu = \frac{\sigma_\alpha}{\sqrt{2}}$ indicates the diversity and σ_α is standard variance of coefficients.

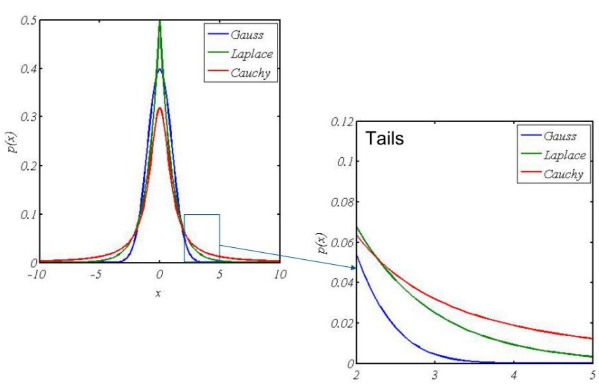

Fig. 1. Some distributions of coefficients.

In contrast to the sharp peak at zero in Laplacian prior, as show in Fig. 1, we can find that Cauchy prior is relatively conservative in the sense of sparseness. In other words, its coefficients is less sparse, a signal is sparse if most entries of the coefficient vector are zero or close to zero. To enforce the prediction accuracy, we employ Cauchy distribution to represent the latent prior in coefficient space, a Cauchy distribution of the variables α_i can be formulated as follows [19] :

$$p(\alpha_i|\sigma_\alpha) \propto \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \alpha_i^2}, \tag{4}$$

the multidimensional distribution is given by $p(\alpha|\sigma_\alpha) = \Pi_i p(\alpha_i|\sigma_\alpha)$, and we have used this distribution as a prior for the vector α. Combining the Cauchy prior, our Cauchy regularized sparse representation model is formulated as follows:

$$J(\alpha) = ||y - D\alpha||_2^2 + \lambda \sum\nolimits_i \ln(1 + \frac{\alpha_i^2}{\sigma_\alpha^2}), \tag{5}$$

where $\sum_i \ln(1 + \frac{\alpha_i^2}{\sigma_\alpha^2})$ is the regularizer imposed by the Cauchy distribution and is a measure of the sparseness of the vector of powers $P_k = \alpha_i^2, i = 0, ..., M-1$. The constant σ_α controls the amount of sparseness.

Taking derivatives of (5) and equating to zero yields

$$\alpha = (D^T D + \lambda Q^{-1})^{-1} D^T y, \tag{6}$$

where λ is a regularization parameter and Q is a $M \times M$ diagonal matrix with elements

$$Q_{ii} = 1 + \frac{\alpha_i^2}{\sigma_\alpha^2}, \qquad i = 0, ..., M-1, \tag{7}$$

The elements of the diagonal matrix Q are nonlinearly related to the weight coefficients α_i. That is Q depend on α, so the equation (6) has to be solved by means of an iterative procedure. The algorithm starts with the coefficient vector $\alpha^{(0)}$. The initial solution is also used to generate the matrix $Q^{(0)}$. In each iteration, we compute

$$Q_{ii}^{iter} = 1 + \frac{(\alpha_i^{iter})^2}{\sigma_\alpha^2}, \tag{8}$$

which we subsequently use to update the coding vector α as

$$\alpha^{iter} = (D^T D + \lambda (Q^{iter})^{-1})^{-1} D^T y, \tag{9}$$

where $iter$ denotes iteration number, then, we use α^{iter} to tune the value of σ_α.

In general, a few iterations (≤ 10) are needed to minimize the cost function $J(\alpha)$.

2.2. Hallucinating face images

Face hallucination refers to the technique of estimating a HR face image from the observed LR face image. Let y be input

LR face image, Y the LR training dictionary whose column vector consists of LR face image $Y_m, m = 1, ...M$, where M is the number of training images. In our method, the reconstruction weights of the each input image patch $y^{(i,j)}$ located at position (i, j) in the LR face image are computed by the following optimization problem:

$$\alpha^{(i,j)*} = \underset{\alpha^{(i,j)}}{\operatorname{argmin}} ||y^{(i,j)} - Y^{(i,j)}\alpha^{(i,j)}||_2^2 + \lambda \sum\nolimits_m \ln(1 + \frac{(\alpha_m^{(i,j)})^2}{\sigma^2_{\alpha^{(i,j)}}}), \tag{10}$$

where $\alpha^{(i,j)}$ represents the linear reconstruction coefficient vector, $Y^{(i,j)}$ are the same position patches in LR training image dictionary. After obtaining the reconstruction coefficients by training LR face images, based on the assumption that LR and HR patch share similar topological manifold structure [2], the coefficients are mapped to HR directly to synthesize the HR face patch $x^{(i,j)}$ through the corresponding HR training dictionary $X^{(i,j)}$ by

$$x^{(i,j)} = X^{(i,j)}\alpha^{(i,j)*}. \tag{11}$$

Consequently, the target HR image x is reconstructed by combining these hallucinated HR patches.

3. EXPERIMENTS AND RESULTS

In this section, we perform experiments on benchmark face databases to demonstrate the performance of our method. Subjective hallucination results and the objective metrics, i.e., PSNR and SSIM indexes, are demonstrated.

3.1. Database Description

Fig. 2. Some training faces in FEI Face Database.

The experiments are conducted on FEI face database [18]. FEI face database composed of only frontal and pre-aligned face images (some samples are shown in Fig. 2). The subset contains 400 images from 200 subjects (100 men and 100 women), those subjects are mainly from 19 and 40 years old with distinct appearances, hairstyles and adornments, and each subject has two frontal images (one with a neutral or non-smiling expression and the other with a smiling facial expression). All the images are cropped to 120× 100 pixels, and we randomly choose 360 images (180 subjects) as the training set, leaving the rest 40 images (20 subjects) for testing. Therefore, all the test images were absent completely in the training set. The LR images are formed by smoothing (an averaging filter of size 4×4) and down-sampling (by a factor of 4) from corresponding HR images. The HR patch size was 12×12 and the overlap between neighbor patches was 4 pixels, while the corresponding LR patch size was 3×3 with an overlap of 1 pixel.

Table 1. PSNR and SSIM comparison of different methods.

methods	PSNR(dB)	SSIM
NE	31.75	0.894
LSR	31.90	0.903
SR	32.11	0.905
NCSR	31.30	0.906
Our method	**32.51**	**0.910**

3.2. Results Comparison

Subjective hallucinated results by different methods such as Chang's NE [2], Ma's LSR [6], Jung's SR [8] and Dong's NCSR [17] as well as our method are demonstrated. The objective metrics, i.e., PSNR and SSIM index, are also compared. For the sake of fair comparison, the control parameters in all methods are tuned to their best results.

Table 1 tabulates the average PSNR (dB) and SSIM [20] comparison of different methods on the 40 testing face images of FEI face database respectively. As shown in Table 1, both PSNR and SSIM values of our method are much higher than those of other methods. Besides the objective comparison, we are more concerned about the subjective visual quality differences. As shown in Fig. 3, the regularized methods, including SR and our method, can generate more competitive results with more facial details than the other two non-regularized methods, i.e., LSR and NE. By a further examining, the hallucinated images of our method are cleaner and more similar to the original HR faces (see the noses and eyes). Compared with that, SR method show their inferiority (smooth the edges and textures), this confirms that a properly chosen regularization can indeed direct the solution toward a better quality outcome. NCSR fails to recover visual details of facial features (see the cheeks and mouths) because it primarily exploits the image nonlocal self-similarity, however, for human face images, the self-similarity assumption does not hold well.

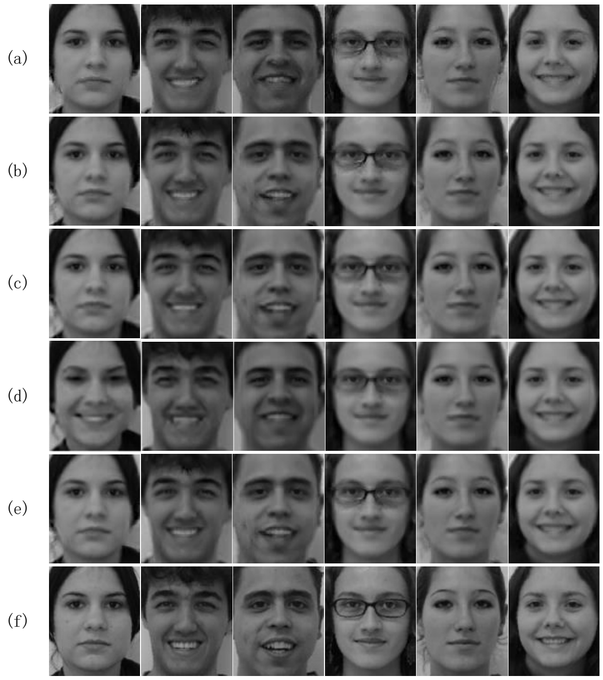

Fig. 3. Subjective results by different methods: (a) NE [2]; (b) LSR [6]; (c) SR [8]; (d) NCSR [17]; (e) our method; (f) original HR faces (ground truth).

4. CONCLUSION

In this paper, we have proposed a moderately sparse prior model to boost the performance of SR based face hallucination. One important advantage of our method lies in its excellent ability to characterize the sparsity of face images with Cauchy regularization term. The plausibility of Cauchy model has been verified by an experiment on benchmark face database, the experimental results clearly demonstrated the superiority in terms of PSNR, SSIM and subjective quality.

5. REFERENCES

[1] S. Baker and T. Kanade, "Hallucinating faces," in *FG*, 2000, pp. 83–88.

[2] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2004, vol. 1, pp. 275–282.

[3] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, vol. 35, no. 3, pp. 425–434, 2005.

[4] B. Li, H. Chang, S. Shan, and X. Chen, "Aligning coupled manifolds for face hallucination," *Signal Processing Letters, IEEE*, vol. 16, no. 11, pp. 957–960, 2009.

[5] X. Ma, J. Zhang, and C. Qi, "Position-based face hallucination method," in *Proc. IEEE Conf. on Multimedia and Expo (ICME)*, 2009, pp. 290–293.

[6] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognition*, vol. 43, no. 6, pp. 2224–2236, 2010.

[7] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face hallucination via sparse coding," in *Proc. IEEE Conf. on Image Processing (ICIP)*, 2008, pp. 1264–1267.

[8] 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, 2011.

[9] Yuan Yuan Xiaoqiang Lu and Pingkun Yan, "Alternatively constrained dictionary learning for image super-resolution," *IEEE Trans. on Cybernetics*, vol. 44, no. 3, pp. 366–377, 2014.

[10] Dacheng Tao Kaibing Zhang, Xinbo Gao and Xuelong Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. on Image Processing*, vol. 21, no. 11, pp. 4544–4556, 2012.

[11] Shenming Qu, Ruimin Hu, Shihong Chen, Liang Chen, and Maosheng Zhang, "Robust face super-resolution via position-patch neighborhood preserving," in *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1–5.

[12] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[13] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[14] Junjun Jiang, Ruimin Hu, Zhen Han, and Tao Lu, “Efficient single image super-resolution via graph-constrained least squares regression,” *Multimedia Tools and Applications*, pp. 1–24, 2013.

[15] Michael B. Wakin Emmanuel J. Candes and Stephen P. Boyd, “Enhancing sparsity by reweighted l1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[16] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, “Image super-resolution via sparse representation,” *IEEE Trans. on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[17] Guangming Shi Weisheng Dong, Lei Zhang and Xin Li, “Nonlocally centralized sparse representation for image restoration,” *IEEE Trans. on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.

[18] C. Thomaz and G. 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.

[19] N. L. Johnson and S. Kotz, “Distributions in statistics: 1 and 2,” *New York: Wiley*, 1970.

[20] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.