FACE HALLUCINATION VIA WEIGHTED SPARSE REPRESENTATION

Zhongyuan Wang, Junjun Jiang, Zixiang Xiong[+]*, Ruimin Hu, Zhenfeng Shao*[*]

NERCMS, School of Computer, [*]LIESMARS ,Wuhan University, China
[+]Texas A&M University, USA

ABSTRACT

By incorporating the priors of image positions, position-patch based face hallucination methods can produce high-quality results and save computation time. These methods represent the test image patch as a linear combination of the same position patches in a training dictionary, and the key issue is how to obtain the optimal coefficients. Due to stability and accuracy issues, methods based on least square estimation or sparse representation (SR) proposed so far are not satisfactory. In this paper, we improve existing SR methods by exploiting similarity between the test and training patches. In particular, we impose a similarity constraint (in terms of the distance between the test patch and bases in the dictionary) on the ℓ_1 minimization regularization term and obtain the coefficients by solving a weighted SR problem. We also provide a new prospective on weighted SR and investigate its robustness to illumination variations. Experiments on commonly used database demonstrate that our method outperforms state of the art.

Index Terms—Super-resolution, face hallucination, weighted sparse representation, position patches.

1. INTRODUCTION

Face super-resolution, or face hallucination, refers to the technique of estimating a high-resolution (HR) face image from low-resolution (LR) face image sequences or a single LR one. Due to constrained imaging conditions in many scenarios, it is hard to capture HR face images, and thus face hallucination is extensively used for pre- and/or post-processing in video applications, such as video surveillance and video retrieval. A large number of theoretical and applied works on face hallucination have been carried out. According to [1], existing methods can be classified into three categories: interpolation, reconstruction-based methods, and machine learning methods. Among them, learning based methods have received much attention because they can achieve high magnifying factors, however, they typically require databases of millions of high- and low-resolution patch pairs, and are therefore computationally intensive.

Following locally linear embedding (LLE) [2] from manifold learning, Chang *et al.* assumed similarity between two manifolds in the HR and LR patch spaces and proposed a neighbor embedding (NE) based face hallucination method [3] using a fixed number of neighbors for reconstruction. Due to under- or over-fitting, this method usually results in blurring. To alleviate this problem, Yang *et al.* [4] employed a sparse coding method to adaptively choose the most relevant neighbors for reconstruction. It should be noted that unlike generic image super-resolution, face hallucination is a specific problem, and therefore image priors can be incorporated to boost its performance. Following this idea, Ma *et al.* [5] took advantage of the prior of face positions and introduced a position-patch based method to hallucinate the HR face image using the same position image patches of all training images. This way, it saves computation time and gives relatively good results. However, when the number of training patches is much larger than the dimension of the patch, least square estimation (LSE) employed to obtain reconstruction coefficients is actually under determined and generates unstable solutions. It is known that ℓ_1 -minimization-based SR can provide stable solutions [6]. Recently, Jung *et al*. [7] formulated a SR version of the position-patch based method and used convex optimization to overcome the problems of LSE. Nevertheless, SR based methods may select very distinct patches that are far from the input patch to favor sparsity and thus result in dissimilarity in terms of the Euclidean distance.

Since the sparse representation theory was established in [6], several variants have been successively developed, namely, weighted ℓ_1 minimization [8-9], SR with prior support information [10], locality-constrained SR [11,13], and structured sparsity [12]. These methods aim at exploiting the *a priori* information about the support of coding coefficients. Candès *et al*. [8] designed an iterative reweighted formulation of ℓ_1 minimization to more democratically penalize nonzero coefficients. Friedlander *et al.* [10] theoretically proved that if the partial support estimate is at least 50% accurate, then weighted ℓ_1 minimization outperforms the standard one in terms of accuracy, stability, and robustness. In practical applications, weighted SR has shown superiority over conventional SR methods for pattern classification applications such as image classification [13-14] and face recognition [15].

Inspired by the above works on reweighted ℓ_1 minimization, in this paper we extend position-patch based face hallucination from SR [7] to weighted sparse representation (WSR) by enforcing a similarity-inducing constraint on the coding coefficients. By incorporating both sparsity and distance support information, WSR can better characterize the similarity between the test sample and training samples and thus give more accurate and robust solutions. Although weighted ℓ_1 -minimization-based sparse representation has been previously studied in [8-15], we are the first to introduce it to the face hallucination problem. Different from [14,15], we study the weighted ℓ_1 minimization problem from a probability theory

The research was supported by the National Natural Science Foundation of China (61231015,61172173,61070080,61003184,61170023), the major national science and technology special projects (2010ZX03004-003-03,2010ZX03004-001-03), the National Basic Research Program of China(973Program) (2009CB320906).

(rather than a penalty) prospective, which can be justified to be preferable theoretically. Furthermore, our employed similarity metric is capable of handling possible illumination mismatch among the input and training images. Experiments on commonly used database confirm superiority of our proposed WSR method.

2. FORMULATIONS

Let X be test 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 position-patch based face hallucination [5], each patch $X(i,j)$ located at position (i,j) in the LR face image can be represented as

$$X(i,j)=Y(i,j)w(i,j) \ , \qquad (1)$$

where $w(i,j)$ represents the linear reconstruction coefficient vector and $Y(i,j)$ are the same position patches in LR training image dictionary.

In [7], the reconstruction weights of the input image patch are computed by the following ℓ_1-norm sparse representation:

$$\min_{w(i,j)} \| w(i,j) \|_1 , \ s.t. \| X(i,j)-Y(i,j)w(i,j) \|_2^2 \le \varepsilon \ , \qquad (2)$$

where $\varepsilon \ge 0$ is the allowed error tolerance. After obtaining the reconstruction coefficients by training LR face images, based on the assumption that LR and HR patch share similar topological manifold structure [3], the coefficients are mapped to HR directly to synthesize the HR face patch $X_H(i,j)$ through the corresponding HR training dictionary $Y_H(i,j)$ by

$$X_H(i,j)=Y_H(i,j)w(i,j) \ . \qquad (3)$$

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

3. WEIGHTED SPARSE REPRESENTATION

For face hallucination based on LR/HR training dictionary pairs, similarity in terms of the Euclidean distance should be more emphasized than correlation [2,11,13]. However, neither ℓ_1-norm nor ℓ_2-norm optimization problem explicitly involves any similarity measure. For example, the commonly used iterative greedy algorithm for an ℓ_1-norm optimization problem finds the index of the single dictionary element that best approximates the current image/residual signal by maximizing the cross-correlation [16]. In addition, the analytical solution for the ℓ_2-norm optimization problem also contains a correlation term [3,13]. Therefore , SR cannot guarantee consistency between coefficient weights and similarity in terms of the Euclidean distance.

To illustrate this issue more clearly, we plot the coding coefficients generated by SR according to the sorted distances in ascending order in Fig. 1.

It is seen that, as the distance increases (or similarity decreases), the coefficient magnitudes exhibit an overall decaying trend but leave many outliers (e.g., those marked in red boxes). This is due to the fact that the SR optimization process only accounts for the correlation but ignores pairwise similarity. Consequently, sparse coding may reconstruct a test patch by training patches that are far from the test sample and thus produce unstable hallucinated results.

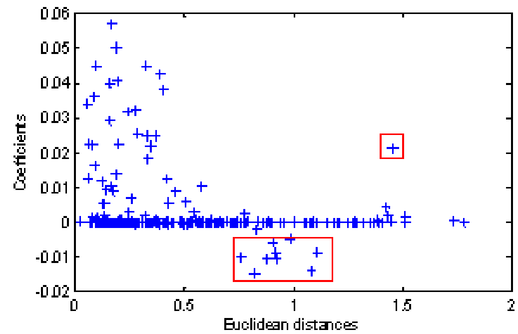


Fig. 1. Sparse coding coefficients vs. Euclidean distances.

3.1 Proposed Method

To overcome the drawbacks of sparse coding, we expect the coefficient magnitudes to be proportional to similarity. From the prospective of probability instead of penalty [13-15], we seek the maximum probability that makes the bases involving significant similarity to retain larger coefficients by

$$w^* = \arg\max_w P(w_i > w_j \mid S_i > S_j), \ \forall \ w_i, w_j \in w, \ S_i, S_j \in S \ , \qquad (4)$$

where w denotes coefficient vector, and S denotes similarity vector which quantitatively expresses the degree of pairwise similarity between the input test patch and each individuals in training dictionary. In statistics, it is a fundamental problem to determine whether two curves, referred as coefficient curve and similarity curve in this paper, share the same variation tendency [17]. In practice, for the sake of simplicity, we maximize the correlation between the coefficient and similarity vectors to obtain $w^* = \arg\max_w (S \bullet |w|)$, where $|w|$ calculates element wise absolute value of the vector w (*i.e.*, magnitude), "•" is the correlation operator. If we use the Euclidean distance d to measure similarity $S=1/d$, then $w^* = \arg\min_w (d \bullet |w|)$, which can be conveniently incorporated into ℓ_1 minimization regularization in (2) because it involves minimization rather than maximization.

Based on the above discussion, our weighted ℓ_1-norm regularization is formulated as

$$w^*(i,j) = \arg\min_{w(i,j)} \left\{ \| X(i,j)-Y(i,j)w(i,j) \|_2^2 + \lambda \| d(i,j) \bullet w(i,j) \|_1 \right\} , \qquad (5)$$

where λ is a regularization parameter balancing the contribution of the reconstruction error and sparsity of the solution, $d(i,j)$ denotes the Euclidean distance vector whose entry $d_m(i,j)$ is calculated as

$$d_m(i,j) = \| X(i,j) - Y^m(i,j) \|_2 \ . \qquad (6)$$

If we use D to denote the diagonal matrix with diagonal elements $d(i,j)$, (5) can be reformulated as

$$w^*(i,j) = \arg\min_{w(i,j)} \left\{ \| X(i,j)-Y(i,j)w(i,j) \|_2^2 + \lambda \| Dw(i,j) \|_1 \right\} \ . \qquad (7)$$

By substituting $w'(i,j) = Dw(i,j)$, $w(i,j) = D^{-1}w'(i,j)$, (7) can be rewritten as

$$w^*(i,j) = \arg\min_{w'(i,j)} \left\{ \| X(i,j)-Y(i,j)D^{-1}w'(i,j) \|_2^2 + \lambda \| w'(i,j) \|_1 \right\} , \qquad (8)$$

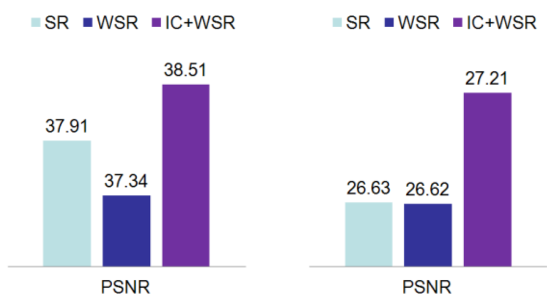
which can be solved using the SLEP toolbox in [18].

3.2 Robustness to Illumination Variations

WSR is proposed to enhance the consistency between coefficient and similarity. However, when the illumination level of test image deviates from that of training images significantly, this enhancement effect is not yet guaranteed. Fig. 2 shows the PSNR of SR as well as WSR for test images with reduced illumination and raised illumination, respectively. It is unexpectedly observed that the PSNR of WSR method is no higher than that of SR any more, especially for reduced illumination case. This phenomenon is attributed to the fact that the standard Euclidean distance cannot accurately measure similarity when the image intensity values do not belong to the same range. So, WSR should account for the illumination intensity variations among test image and training images. Accordingly, we carry out an illumination alignment operation together with the Euclidean distance computation (so-called illumination compensation , IC), and thereby form a weighted Euclidean distance metric which is formulated as

$$g=\sqrt{\frac{(Y^m)^T Y^m}{X^T X}},\quad d_m(i,j)=\| gX(i,j)-Y^m(i,j)\|_2 . \qquad (9)$$

As shown in the third column in Fig. 2 for either (a) or (b) , WSR employing IC (denoted as IC+WSR in Fig. 2) overtakes SR once again, which manifests that our suggested distance metric benefits to WSR in presence of illumination variations. Thus, we use this metric in the subsequent experiments.



(a) Reduced illumination. (b) Raised illumination.

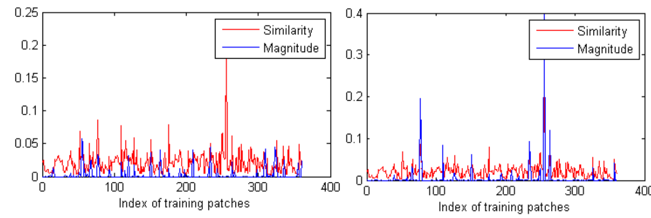
Fig. 2. Results on illumination compensation.

4. EXPERIMENTS AND RESULTS

To verify the superiority of our method, experiments were performed on FEI face database [19]. It contains 400 images from 200 subjects (100 men and 100 women). Among them, 360 images were randomly chosen as the training set, and the rest 40 were used for testing. Therefore, all the test images were absent completely in the training set. The LR images were obtained by smoothing and down-sampling by a factor of 4. The HR patch size was 12x12 and the overlap between neighbor patches was 4 pixels, while the corresponding LR patch size was 3x3 with an overlap of 1 pixel. Since it is already confirmed in [7] that both NE [3] and LSE [5] methods are inferior to SR method, we only compare our method with SR. Subjective hallucination results and the objective metrics, *i.e.*, PSNR and SSIM index, are demonstrated. Both the parameters of the balancing factor λ in WSR and the error tolerance ε in SR are tuned to their best possible results.

4.1 Consistency Between Magnitude and Similarity

The similarity-inducing weighting is designed to seek the most similar variation tendency for similarity and magnitude curves, which can be converted to a maximum cross-correlation problem. To validate this mechanism, we conducted a consistency test, in which the similarity is expressed as inverse distance. As shown in Fig.3 (b), the coefficient curve on WSR coincides with the similarity curve perfectly. In contrast, the curves on SR in Fig.3 (a) perform less similar trend as the large coefficients hardly correspond to the large similarity values. These results confirm that the idea of pursuing the most similar variation tendency for two curves really works.



(a) SR (b) WSR

Fig. 3. Plots of similarity and coefficient magnitude.

4.2 Comparison of Objective and Subjective Results

As shown in Fig. 4, both PSNR and SSIM values of our WSR are much higher than those of SR. For position-patch based face hallucination, three coefficient solving methods are proposed so far, including LSE [5], SR [7] and our WSR .Among them, SR outperforms LSE leading to PSNR gain 0.14dB and SSIM gain 0.0027, respectively. Comparing WSR with SR, it is worth pointing out that PSNR and SSIM gains substantially total up to 0.76dB and 0.0102 . In other words, WSR makes greater progress than SR in terms of PSNR and SSIM, and thus contributes the development of position-patch based face hallucination.

In addition, We chose 5 pictures to show in Fig. 5, and it is obviously observed that the edges and facial details around nose and mouth areas in images generated by WSR are preserved better.

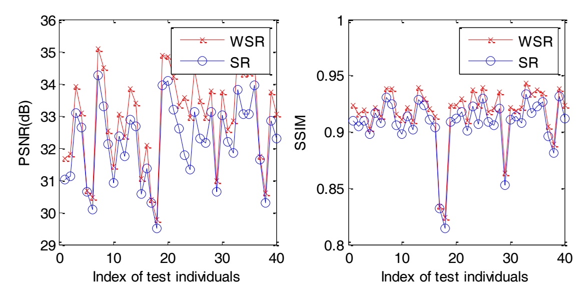


Fig. 4. Objective results on FEI face database. PSNR average gain: 0.76 (WSR 32.93,SR 32.17); SSIM average gain: 0.0102 (WSR 0.9171,SR 0.9069).

To test robustness to noise, we conducted experiments using noise corrupted face images which were added with $\sigma=5$ zero mean Gaussian noises. Some subjective results are shown in Fig.6. We can observe WSR is more robust to noise than SR. WSR almost completely removes interference noises and it does not result in blurring effects on facial features such as mouth, nose and eyes. Meanwhile, results in Fig. 6 are more pronounced than those in Fig. 5. This is partially due to the fact that noisy images are less sparse, and thus sparse representation method can benefit more from similarity constraints in the presence of noise.

Fig. 5. Subjective results on FEI face database. First row: input LR faces (enlarged using bicubic). Second row: SR. Third row: WSR. Fourth row: original HR faces.

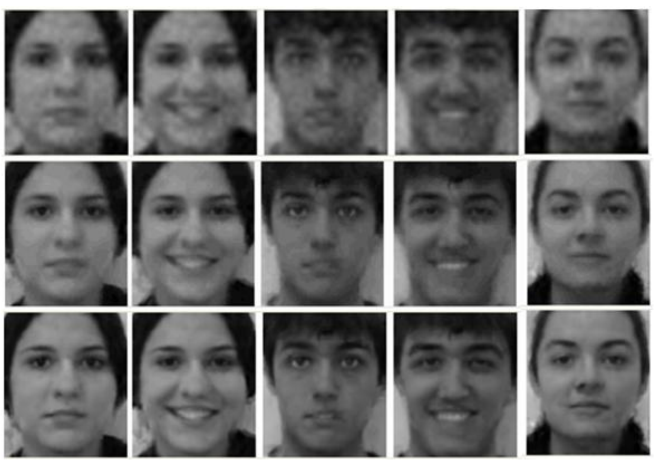


Fig. 6. Results on noises corrupted faces. Top: input LR noisy faces (enlarged using bicubic). Middle: SR. Bottom: WSR.

5. CONCLUSIONS

In this paper, we have proposed a WSR method to boost the performance of position-patch based face hallucination. From the prospective of seeking the most similar variation tendency for similarity and coefficient magnitude series, which actually turns out a maximum cross-correlation problem, we impose a similarity-inducing constraint onto ℓ_1-norm regulation term. Not only can WSR account for correlation but also similarity in terms of distance. This way WSR also avoids training bases that are far from the test patch in the process of face hallucination. Moreover, we suggested a robust distance metric to compensate possible illumination variations present in actual images. Experimental results on FEI face database have demonstrated the superiority of WSR over regular SR in terms of PSNR, SSIM and subjective visual quality. Our future work will focus on extending WSR to the generic image super-resolution problem. In addition, we plan to employ more complex but accurate model, e.g., regression models, to match the variation tendency for coefficient and similarity series in the future.

6. REFERENCES

[1] J.C. Yang, John Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Trans. on Image Processing*, 19 (11):2861-2873, 2010.

[2] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 (5500): 2323–2326, 2000.

[3] H. Chang, D.Y. Yeung, and Y.M. Xiong. Super-resolution through neighbor embedding. In *CVPR*, pp. 275–282, 2004.

[4] J. Yang, H. Tang, Y. Ma, and T. Huang. Face hallucination via sparse coding. *ICIP*, pp. 1264–1267, 2008.

[5] X. Ma, J. Zhang, and C. Qi. Hallucinating face by position-patch. *Pattern Recognition*, 43(6): 3178–3194, 2010.

[6] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Com. Pure Appl. Math.*, vol.59, pp. 1207–1223, 2006.

[7] C. Jung, L. Jiao, B. Liu, and M. Gong. Position-patch based face hallucination using convex optimization. *IEEE Signal Process. Lett.*, 18(6): 367–370, 2011.

[8] E. J. Candès, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877-905, 2008.

[9] S. Oymak, M. A. Khajehnejad, and B. Hassibi .Recovery threshold for optimal weight ℓ1 minimization. *IEEE International Symposium on Information Theory Proceedings*, 2012.

[10] M. P. Friedlander, H. Mansour, R. Saab, and O. Yılmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. on Inf. Theory*, 58(2): 1122-1134, 2012.

[11] K. Yu, T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. In *Proc. of Advances in Neural Information Processing Systems*, pp.2223-2231, 2009.

[12] M. Kowalski and B. Torresani. Structured sparsity: from mixed norms to structured shrinkage. In *Workshop on Signal Processing with Adaptive Sparse Representations*, 2009.

[13] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp.3360–3367, 2010.

[14] J. J. Thiagarajan and Andreas Spanias. Learning dictionaries for local sparse coding in image classification. *Signals, Systems and Computers*, pp.2014-2018 , 2011.

[15] Kui Jia, Tsung Han Chan, and Yi Ma. Robust and practical face recognition via structured sparsity. In *ECCV* , 2012.

[16] S. Mallat and Z. Zhang . Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Process.*, 41(12):3397-3415,1993.

[17] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky. ℓ1 trend filtering. *SIAM Review*, problems and techniques section, 51(2):339–360, 2009.

[18] J. Liu, S. Ji, and J. Ye. SLEP: Sparse learning with efficient projections. Arizona State University, 2010, http://www.public.asu.edu/~jye02/Software/SLEP.

[19] FEI Face Database , http://fei.edu.br/~cet/facedatabase.html.