ROBUST FACE RECOGNITION BASED ON ITERATIVE SPARSE CODING AND PIXEL SELECTION

Lina Lian, Huicheng Zheng and Jiayu Dong

School of Data and Computer Science, Sun Yat-sen University
Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

ABSTRACT

Face recognition based on sparse representation has attracted broad interest in recent years. In many existing sparse coding works, the distribution of error term (coding residual) is modeled with a Laplacian or Gaussian function, which leads to an l_1-norm or l_2-norm minimization problem. However, it is hard to fit the error term satisfactorily in practice, especially when occlusion or corruption exists. In order to improve the robustness of sparse coding algorithms to outliers, we propose an efficient pixel-selection strategy in this paper, which can pick out unspoiled pixels from a seriously damaged face image. However, it is very challenging to determine the positions of the outliers directly. We propose an iterative coding approach to improve the selection. Extensive experiments demonstrate the robustness and effectiveness of the proposed method in the presence of occlusion and corruption.

***Index Terms*—** sparse representation, pixel selection, block occlusion, pixel corruption, face recognition

1. INTRODUCTION

Face recognition is one of the classical and challenging research topics in computer vision, pattern recognition and biometrics research [1]. Many face recognition techniques have been developed over the past several decades [2, 3, 4, 5, 6, 7]. Recently, an interesting work done by Wright et al. [8] showed that sparse representation can be employed for robust pattern recognition with impressive performance. When sufficient training samples are available in each class, a test sample is expected to be best represented as a linear combination of training samples from the same class. Given a test sample $\mathbf{y}$ and a training dictionary A, classification based on sparse representation (SRC, [9]) can be formulated as

$$\hat{\alpha} = \arg\min_{\alpha} \|A\alpha - \mathbf{y}\|_2^2 + \lambda\|\alpha\|_1 \quad (1)$$

This work was supported by National Natural Science Foundation of China (No. 61172141), Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase), Project on the Integration of Industry, Education and Research of Guangdong Province (No. 2013B090500013), Science and Technology Program of Guangzhou (No. 2014J4100092), and Major Projects for the Innovation of Industry and Research of Guangzhou (No. 2014Y2-00213).

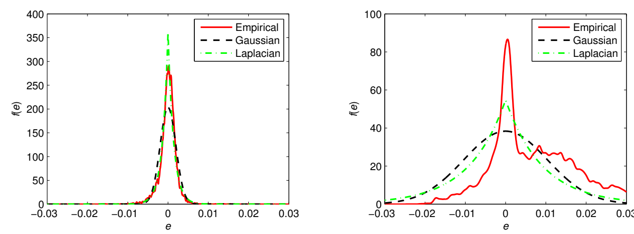

Fig. 1. The empirical and fitted distributions of representation errors for a test image: (a) without occlusion and corruption; (b) with 50% block occlusion plus 50% pixel corruption

where α is the coding vector of $\mathbf{y}$ over A, and λ is a parameter controlling sparsity of the solution. The l_2-norm term in Eq. (1) signifies fidelity or error of the representation, while the l_1-norm term enforces sparsity of coding.

The basic SRC algorithm has achieved encouraging results under controlled conditions, but is not effective for real-world scenarios. Robust error estimation methods [10, 11, 12, 13, 14], which perform error detection or error correction, have been proposed in order to make regression models more robust to outliers (e.g. occlusion and corruption). All these methods assume a priori distribution for the error term. For instance, in [15, 11], the fidelity term is assumed to satisfy a Laplacian or Gaussian distribution by using the l_1-norm or l_2-norm. However, these approaches cannot deal with complex situations with challenging occlusion or corruption (see Fig. 1) [16]. The damaged pixels may introduce great challenges to existing approaches and probably make sparse coding algorithms hard to converge.

In this paper, we propose a new method based on iterative sparse coding and pixel selection, which shows improved performance in dealing with challenging face images subject to heavy occlusion or corruption. We summarize the contributions in this paper as follows:

1. A pixel-selection strategy based on sparse coding is proposed, so that face recognition can be focused on confident pixels instead of outliers for robustness.
2. In view of the difficulty in determining the exact out-

liers, an iterative coding approach is introduced to optimize the selection of pixels based on initial estimations.

3. With pixel selection, only confident pixels in the images are preserved for recognition, which effectively cuts down the dimensions of feature vectors and reduces the computational complexity.

4. The proposed method can also be combined with existing algorithms to improve their performance. Experimental results verified the effectiveness of our idea.

The rest of this paper is organized as follows. Section 2 discusses the limitations of classical sparse coding methods, and then introduces our solution and the corresponding algorithm. We present extensive experimental results and insightful analysis in Section 3. Finally, this paper is concluded by Section 4.

Algorithm 1 : Iterative Sparse Coding and Pixel Selection

Input: Test image $\mathbf{y}$, dictionary A with n training images
Output: Identity ID($\mathbf{y}$) of the test image $\mathbf{y}$

$t = 0$, initialize $\mathbf{y}_{\text{rec}}^t$ as the mean of all training images, $P^t = \mathbf{I}$.
repeat
  $\mathbf{e}^t = \mathbf{y} - \mathbf{y}_{\text{rec}}^t$.
  Compute τ^t with Eq. (4).
  Compute P^{t+1} with Eq. (3).
  $\Psi^{t+1} = \{\, i \,|\, p_i^{t+1} = 1,\ i \in \{1, 2, \cdots, m\}\}$.
  Construct $\tilde{P}^{t+1}$.
  Determine α^{t+1} by solving Eq. (2).
  $\mathbf{y}_{\text{rec}}^{t+1} = A\alpha^{t+1}$.
  $t = t + 1$.
until Convergence or the maximal number of iterations.
$\text{ID}(\mathbf{y}) = \arg\min_d \left\| \tilde{P}^t\mathbf{y} - \tilde{P}^t A\delta_d(\alpha^t) \right\|_2$.

2. THE PROPOSED METHOD

In this section, we analyze the limitations of existing sparse coding strategies and propose an iterative pixel-selection approach for improved robustness in sparse coding.

2.1. Error Estimation in Classical Sparse Coding Methods

Defining the error term with l_1-norm or l_2-norm actually assumes that the coding residual **e** follows a Laplacian or Gaussian distribution. However, this assumption may be far from the truth, especially when occlusion, corruption or other variations take place [16].

To verify our conjecture, two controlled experiments are carried out on the EYB database [6, 7] by following the experimental setting in Section 3.1. One of the test samples is free of any occlusion and corruption, while the other one suffers from 50% block occlusion plus 50% random pixel corruption. The empirical distribution of sparse coding errors as well as those fitted by using a Laplacian or Gaussian function are shown in Fig. 1. The difference between Fig. 1(a) and Fig. 1(b) shows that error estimation in classical sparse coding methods may be suitable for face images without occlusion and corruption, but could be far from the actual distribution under challenging situations when heavy occlusion or corruption takes place. In fact, the errors due to outliers may follow arbitrary and unexpected distributions. Therefore, it may be inappropriate to impose hand-crafted distribution models for the errors. Instead, we propose to remove pixels likely corresponding to outliers by using an iterative pixel selection procedure, as introduced in the following.

2.2. Pixel-Selection Sparse Coding

In the proposed method, we aim to select only clean pixels from the contaminated test image for face recognition. A novel objective function is accordingly proposed as follows:

$$\hat{\alpha} = \arg\min_{\alpha} \left\| \tilde{P}A\alpha - \tilde{P}\mathbf{y} \right\|_1 + \lambda\|\alpha\|_1, \qquad (2)$$

where $\tilde{P}$ indicates a *selection* matrix. Compared to Eq. (1), our method aims to select and preserve those pixels which are conducive to face identification, while abandoning those considered as damaged pixels. In other words, $\tilde{P}$ is a projection matrix that transform the original representation into a low-dimension space (with the possible outliers removed).

To calculate the matrix $\tilde{P}$, we define a diagonal matrix $P = diag(p_1, p_2, \cdots, p_m)$ first, where

$$p_i = \begin{cases} 0 & , |e_i| \geq \tau \\ 1 & , |e_i| < \tau \end{cases} \qquad s.t. \quad \mathbf{e} = \mathbf{y} - A\alpha. \qquad (3)$$

In Eq. (3), $\mathbf{e} = [e_1, e_2, \cdots, e_m]$ corresponds to errors in the face image representation and τ is a threshold. For determination of threshold τ, we intuitively take the average error as the default choice in this paper, i.e.,

$$\tau = \frac{1}{m}\sum_i |e_i| \qquad (4)$$

Define $\Psi = \{\, i \,|\, p_i = 1,\ i \in \{1, 2, \cdots, m\}\}$ as the set of locations of the pixels whose reconstruction errors are lower than τ. Let c ($c<m$) be the number of elements in Ψ and $\Psi = \{\psi_1, \psi_2, \cdots, \psi_c\}$. Once P is obtained, the projection matrix $\tilde{P}$ is constructed by selecting the c rows in P corresponding to the locations in Ψ.

2.3. Iterative Optimization

In practice, it is often challenging to determine the exact outliers at the very beginning. So we propose an iterative method to optimize the discrimination of outliers from initial estimations. At each iteration t, we evaluate the reconstruction error

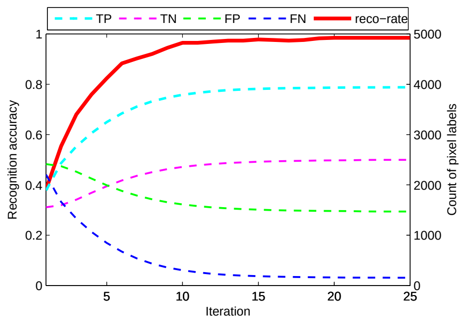

Fig. 2. The effect of iterative sparse coding and pixel selection. TP indicates true positive face pixels, while TN = True Negative, FP = False Positive, and FN = False Negative.

(a) original (b) test (c) iter=1 (d) iter=5 (e) iter=10 (f) iter=20

Fig. 3. Illustration of iterative pixel selection by the proposed method. The top row shows a face image from the EYB database and the bottom row is a face image from the AR database. (a) The original test images; (b) Test images subject to 50% block occlusion or real disguise and 50% pixel corruption; (c)-(f) The detected outliers (shown in black) and the selected pixels (shown in white) varying with iterations.

$\mathbf{e}^t$ of the face image, and then determine the threshold τ^t with Eq. (4). The projection matrix $\tilde{P}^{t+1}$ is constructed accordingly to discard pixels corresponding to outliers estimated at the t-th iteration. Next, α^{t+1} is updated based on pixels considered reliable by solving Eq. (2), which is further used to update the reconstructed image $\mathbf{y}_{\text{rec}}^{t+1} = A\alpha^{t+1}$. The iterations repeat until convergence or a maximal number of iterations is reached.

With the iteratively-updated selection matrix $\tilde{P}$ and the corresponding coding vector α, the identity of a face image can be determined by selecting the class that yields the minimum reconstruction error as follows,

$$\text{ID}(\mathbf{y}) = \arg\min_d \left\| \tilde{P}\mathbf{y} - \tilde{P}A\delta_d(\alpha) \right\|_2 \qquad (5)$$

where $\delta_d(\alpha) \in R^n$ is a vector whose only nonzero components are the elements of α corresponding to the d-th class. The overall method is summarized in Algorithm 1.

Fig. 2 illustrates the effectiveness of the proposed iterative sparse coding and pixel selection. The experiment follows the setting in Section 3.1, in which the level of block occlusion is 50%. TP refers to good pixels being correctly preserved, while TN refers to damaged pixels being correctly discarded. FP or FN indicates pixels that are incorrectly labeled as good ones or bad ones. At each iteration, TP, TN, FP, and FN are averaged over the whole test set and recorded. As we can see from Fig. 2, the amount of true labels (TP + TN) increases and that of false labels (FP + FN) decreases with the iterations (right vertical axis), leading to improved accuracy for face recognition (left vertical axis).

3. EXPERIMENTS

In this section, we present experimental results to compare the proposed method with other related sparse coding algorithms, such as S-SRC [9], R-SRC [9], HQ-A [13], HQ-M [13], RRC-L1 [11], RRC-L2 [11], and CESR [14]. The experiments are carried out on two public face databases: the EYB database [6, 7] and the AR database [17]. Inspired by [18, 10], we design two different experiments to compare the robustness of various sparse coding methods against contiguous block occlusion and random pixel corruption.

In all our experiments, we set $\lambda = 0.2$ empirically. The maximum number of iterations is set as 25 to guarantee reliable convergence. For fairness, all the results presented by competing methods are obtained through identical experiment with optimal parameter.

3.1. Recognition under Occlusion

The first experiment is carried out on the EYB database, which contains 2414 images of 38 subjects. The face images are aligned and cropped into 96×84 pixels as in [9, 11, 13]. Subsets 1 and 2 are chosen for training (717 face images in total) and Subset 3 for testing (453 face images). A square block of a certain size on each test image is replaced by a baboon image as in Fig. 3(b). The location of occlusion is random for each test image and unknown to recognition algorithms.

Fig. 4 shows the recognition rates under various levels of occlusion ranging from 0% to 90%. Our method improves the accuracy over state-of-the-art sparse coding methods, especially when the occlusion level is around 50% $\sim$ 60%, which is very challenging and practical. When half of the test image is occluded, our method still has a recognition accuracy close to 100%, significantly higher than 89.2% of HQ-M, 87.9% of HQ-A, 83.9% of RRC-L1, and 83.4% of RRC-L2.

For more insightful understanding, a test face image subject to 50% block occlusion is used to illustrate how the iterative pixel-selection strategy works, as shown in the first row of Fig. 3. From the results, we can see that with the iterations going on, more and more occluded pixels tend to be determined as outliers while more true face pixels are selected for sparse coding, which is beneficial for robustness of face recognition.

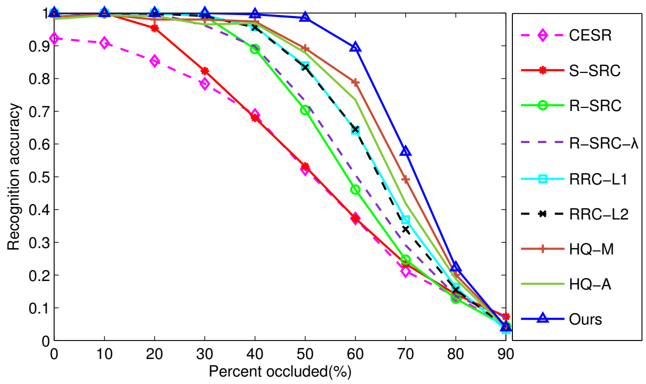

Fig. 4. Face recognition rates on the EYB database under various levels of block occlusion.

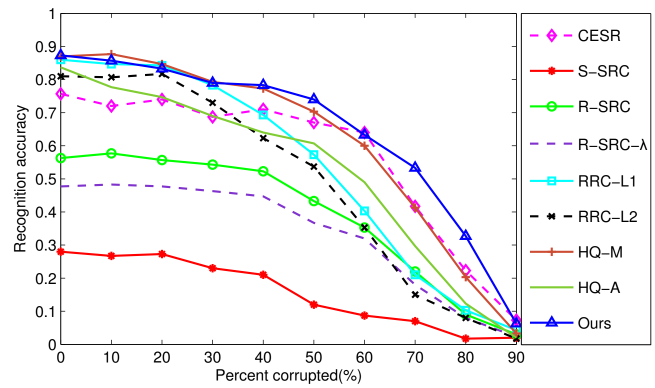

Fig. 5. Face recognition rates on the AR database under real disguise and various levels of random pixel corruption.

3.2. Recognition under Real Disguise and Corruption

The second experiment is carried out on the AR database, which contains over 3000 frontal images of 126 subjects. 100 subjects (50 males and 50 females) are chosen randomly from the database. For each subject, 4 images of non-occluded frontal view from Session 1 are used for training and the other 3 images containing illumination variation and facial disguise (wearing sunglasses) from Session 2 are used for testing. A certain percentage of randomly chosen pixels in each test image are corrupted by replacing those pixel values with random values independently drawn from a uniform distribution in [0, 255]. All the images were resized to 42×30 pixels.

In the experiment, the level of random pixel corruption varies from 0% to 90%. The corresponding results are summarized in Fig. 5. Our method demonstrates superiority to other existing methods in most cases, especially when the conditions get challenging. The performance of several other methods like RRC-L1, RRC-L2, and HQ-A is satisfactory when pixel corruption is sparse, but drops abruptly with increasing percentage of random corruption. For 70% and 80% random corruption, our method shows an improvement of accuracy by more than 12% compared to the best results reported by state-of-the-art methods. The experimental results verified that our method is very robust agaist the combination of sunglass occlusion and dense pixel corruption.

The second row of Fig. 3 illustrates the iterative pixel-selection process for a test image under real disguise and 50% corruption. Our method basically removes most of the outliers corresponding to the sunglasses and corruption while preserving true face pixels, which verified its effectiveness.

3.3. Improving Existing Methods

Existing sparse coding methods often show poor performance on face images subject to substantial occlusion or corruption. In this section, we apply the proposed pixel-selection method to several existing algorithms to improve their robustness. As representative examples, we apply the proposed method to RRC-L1, RRC-L2, and S-SRC, and then carry out comparative experiments under various levels of block occlusion.

Table 1. Comparison of recognition accuracies on the EYB database under various levels of block occlusion.

	30%	40%	50%	60%	70%	80%
RRC-L1	0.993	0.958	0.839	0.642	0.369	0.163
RRC-L2	0.991	0.956	0.834	0.645	0.340	0.155
S-SRC	0.823	0.680	0.532	0.373	0.234	0.141
Ours+RRC-L1	0.993	0.989	0.956	0.859	0.576	0.249
Ours+RRC-L2	0.996	0.982	0.958	0.865	0.570	0.258
Ours+S-SRC	0.985	0.969	0.943	0.817	0.528	0.252

Table 1 shows the comparative results before and after applying the proposed iterative pixel-selection strategy. The performance of existing methods is significantly improved by the proposed method. For example, the original RRC-L1, RRC-L2, and S-SRC achieve 36.9%, 34.0%, and 23.4% recognition accuracies, respectively, when the level of occlusion is 70%. When combined with the proposed iterative pixel-selection strategy, the accuracies have all been improved by more than 20%.

4. CONCLUSION

In this paper, we present an iterative pixel-selection strategy to improve the robustness of face recognition against contiguous occlusion and discontiguous corruption. Instead of fitting the coding errors with probabilistic distributions of fixed forms, we propose to remove likely outliers with a pixel-selection matrix. An iterative procedure is introduced to improve selection of pixels starting from an initial estimation. The proposed method can be easily applied to existing algorithms to improve their robustness. There are still few interesting directions to be explored, such as spatial correlation between pixels or detailed threshold analysis.

5. REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[2] A. Hadid T. Ahonen and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[3] Y. Sun, X. Wang, and X. Tang, "Deeply learned face recognition are sparse, selective and robust," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2892–2900, Jun. 2015.

[4] I. Fedorov, R. Giri, B.D. Rao, and T.Q. Nguyen, "Robust Bayesian method for simultaneous block sparse signal recovery with applications to face recognition," *IEEE International Conference on Image Processing*, pp. 3872–3876, Sep. 2016.

[5] M. Johnson and A. Savakis, "L1-grassmann manifolds for robust face recognition," *IEEE International Conference on Image Processing*, pp. 482–486, Sep. 2015.

[6] K.C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, May 2005.

[7] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[8] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceeding of the IEEE*, vol. 98, no. 7, pp. 1031–1044, 2010.

[9] J. Wright, A. Yang, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[10] M. Iliadis, L. Spinoulas, A.S. Berahas, H. Wang, and A.K. Katsaggelos, "Multi-model robust error correction for face recognition," *IEEE International Conference on Image Processing*, pp. 3229–3233, Sep. 2016.

[11] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Transaction on Image Processing*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[12] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 625–632, Jun. 2011.

[13] R. He, W.S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 261–275, Feb. 2014.

[14] R. He, W.S. Zheng, , and B. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[15] M. Yan, D. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 625–632, Jun. 2011.

[16] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 156–171, Jan. 2017.

[17] A.M. Martinez and R. Benavente, "The AR face database," in *Technical Report #24*. Centre de Visiò per Computador, Universitat Autònoma de Barcelona, 1998.

[18] J. Zheng, P. Yang, S. Chen, G. Shen, and W. Wang, "Iterative re-constrained group sparse face recognition with adaptive weights learning," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2408–2423, May 2017.