The Realization of Face Recognition Algorithm Based on Compressed Sensing (Short Paper)

Huimin Zhang[1,2](✉), Yan Sun[3], Haiwei Sun[3], and Xin Yuan[3]

[1] University of Michigan, Ann Arbor, MI 48105, USA
lizzyww@umich.edu
[2] Shanghai Jiao Tong University, Shanghai 200240, China
[3] Jiangsu University, Zhenjiang 212013, Jiangsu, China
614564957@qq.com, 329788350@qq.com, jund85@163.com

Abstract. Once the sparse representation-based classifier (SRC) was raised, it achieved a more outstanding performance than typical classification algorithm. Normally, SRC algorithm adopts l_1-norm minimization method to solve the sparse vector, and its computation complexity increases correspondingly. In this paper, we put forward a compressed sensing reconstruction algorithm based on residuals. This algorithm utilizes the local sparsity within figures as well as the non-local similarity among figure blocks to boost the performance of the reconstruction algorithm while remaining a median computation complexity. It achieves a superior recognition rate in the experiments of Yale facial database.

Keywords: Compressed sensing · Face recognition · Feature extraction · Sparse representation classification · Image reconstruction

1 Introduction

Face recognition technology, as one of the most significant and successful applications of image analysis and understanding, has now become a research focus in computer vision field. A potential issue one may face is that some features may be lost during data collection and compression. To solve this problem, Donoho and Cande et al. [1] put forward a compressed sensing algorithm that combined sampling and compression. They performed a non-adaptive linear projection to input signal sampling, and reconstructed the original signal using corresponding reconstruction algorithms. This method helped decrease the size of data as well as offered possibility to realize a better restoration even in the case of small sampling data. Based on this method, our paper raised a compressed sensing reconstruction algorithm based on residuals to deal with the large data volume in image signal processing.

H. Gao et al. (Eds.): CollaborateCom 2018, LNICST 268, pp. 571–580, 2019.
https://doi.org/10.1007/978-3-030-12981-1_40

2 Sparse Representation-Based Classifier

Sparse representation-based classifier (SRC) theorem supposes that each class of testing sample has enough training samples. Then, the testing data can be linear represented by training data, and other categories of samples has zero contribution to the reconstruction of this testing samples. Therefore, a common signal classification problem has been transformed into a sparse representation problem. Through computing the category of minimal reconstruction error between testing sample and training sample, we are able to judge the class of testing figures.

The mathematics model of SRC algorithm is as follow:

Suppose n training sample belongs to c classes separately, namely $\{(b_i, g_i)|(b_i \in \chi \subseteq R^m, g_i \in \{1, 2, ..., c\}, i = 1, 2, ..., n\}$ where g_i is the classification label of b_i, m is the dimension of the space, χ and c is the number of all categories. Suppose the alignment of i^{th} class of training sample is the column of matrix $D_i = \{b_{i,1}, b_{i,2}, ..., b_{i,n_i}\} \in R^{m \times n_i}, i = 1, 2, ...c$, where $b_{i,j}$ represents samples in i^{th} class and n_i is the number of training samples in i^{th} category. Then the dictionary matrix can be represented as:

$$D = \{D_1, D_2, ...D_c\} \in R^{m \times n} \tag{1}$$

where $n = \sum_{i=1}^{c} n_i$.

According to SRC theorem, the testing sample can be linearly represented by the dictionary atoms as:

$$b = D\alpha \tag{2}$$

where $\alpha \in R^n$ represents the coefficient vector of atoms in the linear combination.

Suppose the testing sample b belongs to j^{th} class, then the elements in coefficient vector α are all zero except that part relevant to j^{th} class, which can be written as:

$$\alpha = [0, ..., 0, \alpha_{j,1}, ..., \alpha_{j,n_j}, 0, ..., 0]^T \tag{3}$$

where $\alpha_{j,i} \in R$ is the corresponding coefficient of training sample $b_{j,i}$.

Consequently, coefficient α is sparse, and the testing sample b can be represented by linear combination of training data in j^{th} class. The problems of seeking for sparse coefficient vector α can be represented as:

$$\underset{\alpha}{min} \|\alpha\|_0 \quad s.t. \quad b = D\alpha \tag{4}$$

This is a NP-hard problem, and one may only attain an approximate solution efficiently. The problems in Eq. (4) can be approximated as:

$$\underset{\alpha}{min} \|\alpha\|_1 \quad s.t. \quad b = D\alpha \tag{5}$$

where $\|\alpha\|_1$ denotes l_1-norm. Take environmental noise into consideration, Eq. (5) can be transformed as:

$$\min_{\alpha} \|\alpha\|_1 \quad s.t. \quad \|b - D\alpha\|_2 \leq \epsilon \tag{6}$$

For each class i, suppose δ_i is the eigenfunction that selects the coefficient associated with i^{th} category, which is defined as:

$$\delta_i(\alpha_j) = \alpha_j, \quad if \quad g_j = i \tag{7}$$

By exacting the coefficient containing the information of i^{th} class, δ_i can be represented by a new vector:

$$\delta_i = [\delta_i(\alpha_1), \delta_i(\alpha_2), ..., \delta_i(\alpha_n)]^T \tag{8}$$

Finally, by minimizing the reconstruction error, we are able to obtain the class that testing sample b belongs to:

$$identity(b) = \underset{i=1,\ldots,c}{argmin}\, r_i(b) = \|b - D_i\delta_i\|_2 \tag{9}$$

3 Face Recognition Algorithm Based on Compressed Sensing

3.1 Compressed Sensing Reconstruction Algorithm Based on Residuals

Compressed sensing algorithm performed a linear projecting of signal $x \in R^N$ to a $M \times N$ measurement matrix ϕ , $M \ll N$,

$$y = \phi x \tag{10}$$

where $y \in R^M$ is a measurement vector. The obtained subrate can be defined as M/N, and the sparse matrix is named ψ. To solve the reconstruction, one need to solve the following optimization problem:

$$\hat{x} = \underset{x \in R^N}{argmin}\, \|\psi x\|_1 \quad s.t. \quad y = \phi x \tag{11}$$

According the Restricted Isometry Property (RIP), if M is big enough, ϕ and ψ would satisfy the incoherence and isometric requirements in compressed sensing reconstruction.

Documents [2,3] have made tremendous efforts in reducing the computation complexity in compressed sensing reconstruction algorithm as well as improving the quality of image reconstruction. However, due to the multidimensional characteristics of image or video data, these two reconstruction methods exert heavy load on storage and computation. Besides, measurement matrix ϕ is large, and requires a larger storage space. To alleviate these problems, document [4] raises the block compressed sensing (BCS) method, in which images are divided into $\sqrt{N_a} \times \sqrt{N_a}$ size non-overlapping blocks, and can realize block-by-block sensing independently. That means $y_i = \Phi_a x_i$, where x_i is i^{th} image block, and Φ_a is a $M_a \times N_a$ measurement matrix. Document [2] raises the BCS-SPL algorithm. It is realized through doing image reconstruction to each image block:

$$\hat{x_i} = BCS - SPL(y_i, \Phi_a) \tag{12}$$

where $BCS - SPL(\cdot)$ is a smooth projection Landweber compressed sensing reconstruction algorithm based on blocking, which ensures the local sparsity. Document [5] suggests adding image block prediction into compressed sensing reconstruction algorithm. It enhances the reconstruction quality through implicitly increasing the compressibility of signals, and will not make any changes to signal sampling process. In this algorithm, the prediction of current image block is based on adjacent image blocks during reconstruction process.

Suppose existing an image block x_i and its measured value $y_i = \Phi_a x_i$. $\tilde{x_i}$ is the prediction of x_i derived from its adjacent image blocks. Then the residual in measurement domain can be derived from $r_i = y_i - \Phi_a \tilde{x_i}$. The final reconstruction calculation method is expressed as:

$$\hat{x_i} = \tilde{x_i} + BCS - SPL(r_i, \Phi_a) \tag{13}$$

Since residual (the difference between x_i and $\tilde{x_i}$) is more compressible (sparse) than image block itself, we can reconstruct a better approximated signal through compressed sensing theory.

Our paper will depict how to apply sparse representation in over-complete dictionary to the algorithm efficiently. Suppose there is a image X with size $\sqrt{N} \times \sqrt{N}$, and it is divided into K pieces of $\sqrt{N_s} \times \sqrt{N_s}$ non-overlapping blocks, and realizes block-by-block sensing independently. From mathematical perspective,

$$y_i = \Phi_s x_i (i = 1, 2, ...K) \tag{14}$$

where $x_i \in R^{N_s}$ is the vector representation of i^{th} image block, $y_i \in R^{M_s}$ is a measurement vector and Φ_s is a $M_s \times N_s$ measurement matrix. Notably, all the image blocks were sampled using the same sub-rate, meaning the global sampling rate is M_s/N_s.

If one try to recover the whole image X through reconstructing image blocks given measurement value y_i, the first thing is to seek out a best prediction. One method to settle this problem is solving the following minimization problem by using some image prior knowledge:

$$\hat{X} = \underset{P \in R^N}{argmin} \|X - P\|_2 + \lambda \Theta(P) \tag{15}$$

where $\|X - P\|_2$ is l_2-norm data fidelity term, $\Theta(P)$ represents the regularized item of the prior image, and λ is the normalized parameter.

The proposed algorithm did not seek out proper prediction image P in Eq. (15) directly. Instead, it realizes the sparse representation of image blocks through improving local sparsity and non-local self-similarity of image blocks, and therefore finds out a best prediction. In short, it makes prediction to figures based on group sparse representation.

Suppose image P is divided into H overlapping image blocks with size $\sqrt{N_p} \times \sqrt{N_p}$. Each small block is represented by vector $p_i \in R^{N_p}(i = 1, 2, ...H)$. Then pile up the best matching blocks p_i in $W \times W$ window to form a $N_p \times C$ matrix based on Euclidean distance. Denote it as $B_i \in R^{N_p \times C}$. Each matrix $B_i(i = 1, 2, ...H)$ contains all the image blocks possessing similar structures. Define the operator to extract matrix B_i from P as $G_i(\cdot)$. $B_i = G_i(P)$. Since the image blocks are overlapped, one can restore image P from B_i as:

$$P = \sum_{i=1}^{H} G_i^T(B_i) \cdot / \sum_{i=1}^{H} G_i^T(E_{N_p \times C}) \tag{16}$$

in which $G_i^T(\cdot)$ is the operator to put each small block B_i back to reconstructed image, otherwise it is zero. The operator $\cdot/$ means dividing element by element. $E_{N_p \times C}$ is an identity matrix with size $N_p \times C$. Equation (16) shows that the reconstructed figures are derived from taking the average value of all the overlapped blocks. Using overlapped blocks can ensure that the obtained prediction will not be affected by the blocking artifacts, and therefore improves the prediction quality of images.

Suppose each group of similar blocks B_i can be linearly represented by some atoms from over-complete dictionary D. Notably, the dictionary structure based on sparse group representation differs from the dictionary form based on sparse block representation. Suppose $D = [d_1, d_2, ...d_M]$. Each atom d_j is a $(N_p \times C) \times M$ matrix, and have the same size as B_i. Given dictionary D with M atoms, seeking out the sparse representation of each B_i for D can be equate to finding sparse vector $\hat{\alpha}_i \in R^M$, that is $\hat{B}_i \simeq \sum_{j=1}^{M} \hat{\alpha}_{i,j} d_j$ where $\hat{\alpha}_{i,j}$ is the j^{th} element in sparse vector α_i. Then, the sparse coding of B_i in dictionary D is:

$$\hat{\alpha} = \underset{\alpha_i \in R^M}{argmin} \|B_i - \sum_{j=1}^{M} \alpha_{i,j} d_j\|_2 + \lambda \|\alpha_i\|_0 \tag{17}$$

where λ is the regularization parameter. The sparsity is measured by l_0-norm of α_i, and the definition of B_i has taken account of the similarity of image blocks. Based on Eq. 16, reconstructing the whole image P from sparse vector $\alpha_i(i = 1, ..., H)$ can be represented as;

$$P \simeq D \odot \hat{\alpha} = \sum_{i=1}^{H} G_i^T(\hat{B}_i) \cdot / \sum_{i=1}^{H} G_i^T(E_{N_p \times C}) \tag{18}$$

where $\hat{\alpha}$ represents the cascade of all $\hat{\alpha}_i$, and the symbol $D \odot \hat{\alpha}$ is used to simplify following equations.

Based on Eqs. (15) and (17), the image prediction based on sparse representation can be expressed as: $\tilde{X} = D \cdot \tilde{\alpha}$. Therefore,

$$\tilde{\alpha} = \underset{\alpha}{argmin} \|X - D \odot \alpha\|_2 + \lambda \|\alpha\|_0 \tag{19}$$

Since X is unknown during the compressed sensing reconstruction procedure, one may not obtain the solution of Eq. (19). This paper uses iterative process to generate best prediction image. Each iteration includes two stages, the first stage is to solve Eq. (20):

$$\tilde{\alpha}^{[k]} = \underset{\alpha}{argmin} \|\tilde{X}^{[k]} - D \odot \alpha^{[k]}\|_2 + \lambda \|\alpha^{[k]}\|_0 \tag{20}$$

Since l_0 is a non-convex optimization, finding solution with respect to Eq. (20) is NP-hard. This problem can be solved by method provided in document [6]. Therefore, the prediction obtained in k^{th} iteration can be represented as $\tilde{X^{[k]}} = D \odot \tilde{\alpha}^{[k]}$.

The second stage employs residual reconstruction algorithm to recover the image. Divide the predicted image $\tilde{X^{[k]}}$ into K non-overlapping blocks $\tilde{x_i}^{[k]}(i = 1, 2, ..., K)$ with size $\sqrt{N_s} \times \sqrt{N_s}$. Use $\tilde{y_i}^{[k]} = \Phi_s \tilde{x_i}^{[k]}$ to perform block sampling, and use $r_i^{[k]} = y_i^{[k]} - \tilde{y_i}^{[k]}$ to calculate reconstruction residual. Then, the computational method of recovering image block is expresses as:

$$\hat{x_i}^{[k]} = \tilde{x_i}^{[k]} + BCS - SPL(r_i^{[k]}, \Phi_s) \tag{21}$$

Finally, merging all the reconstructed non-overlapping blocks to obtain the reconstructed image $X^{[\hat{k+1}]}$. Improve the reconstruction quality through iterative generating prediction and refactoring residuals.

The basic procedures of compressed sensing reconstruction algorithm based on residuals are as follow:

Input: measurement matrix Φ_s, measurement value $y_i(i = 1, 2, ...K)$, window size W based on best Euclidean distance matching, regularization parameter λ, overlapping block size N_p and non-overlapping block size N_s.

Output: reconstructed image $\hat{X}$.

Initialization: $k = 0$ and $\hat{X}^{[0]} = BCS - SPL(y_i^{[k]}, \Phi_s)$

Repeat the following five steps:

Step 1: Solve $\tilde{\alpha}^{[k]} = \underset{\alpha}{argmin} \|\tilde{X}^{[k]} - D \odot \alpha^{[k]}\|_2 + \lambda \|\alpha^{[k]}\|_0$.

Step 2: Calculate the predicted image $\tilde{X^{[k]}} = D \odot \tilde{\alpha}^{[k]}$.

Step 3: Divide $\tilde{X^{[k]}}$ into K non-overlapping image blocks $[\tilde{x_i}^{[k]}]_{i=1}^{K}$.

Step 4: From $i = 1$ to K, calculate both $r_i^{[k]} = y_i^{[k]} - \Phi_s \tilde{x_i}^{[k]}(i = 1, 2, ...K)$ and $\hat{x}_i^{[k]} = \tilde{x}_i^{[k]} + BCS - SPL(r_i^{[k]}, \Phi_s)$.

Step 5: Place $[\hat{x}_i^{[k]}]_{i=1}^{K}$ back to the whole image plane $X^{[\hat{k+1}]}$

Stop iteration until $\|\hat{X}^{[k+1]} - \hat{X}^{[k]}\|_2 \leq 0.001$.

3.2 Improved Uniformly Blocking SRC Algorithm

During the face recognition procedure, dividing the image into blocks to reconstruct and solve sparse vectors can efficiently reduce the computation complexity. The idea of uniform partitioning is to divide facial image into 2×2, 3×3 and 4×4 image blocks, as is shown in Fig. 1.

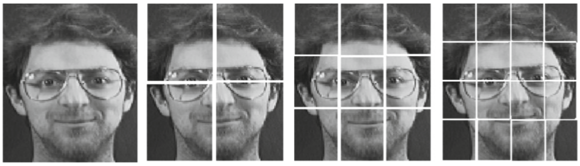

Fig. 1. Uniformly blocking figure

Firstly, uniformly blocking SRC algorithm divides the facial image into K image blocks with same size. Then, it performs SRC algorithm to each image block, and gets the category of each image block with minimal reconstruction error. Finally, the category of all the image blocks is based on majority vote; therefore one can infer the category of the whole facial image.

Combing with the compressed sensing reconstruction algorithm based on residual, it optimizes the uniformly blocking SRC algorithm in solving sparse vector. The procedures of improved uniformly blocking SRC face recognition algorithm are given below.

(1) Input n training facial figures with c classes.
(2) Divide each training figure into q blocks with equal size according to a certain method. Then, transform each image block to the corresponding column vector to form the sub-sample matrix A_i, i=1,2,...q.
(3) Divide the testing facial figure t according to the same method. Then, transform each figure block into column vector to get sub-testing facial figure blocks t_i, i=1,2,...q.
(4) For each sub-block t_i, choose its corresponding sub-sample matrix A_i, and gain the reconstructed image blocks by solving corresponding sparse vector. Use the compressed sensing reconstruction algorithm based on residuals.
(5) Calculate the reconstruction error between reconstructed image blocks and original testing image blocks, and categorize each image block based on minimal reconstruction error.
(6) The category of the whole image is based on majority principle. Ideally, the majority category of image blocks is the actual category among q image blocks. Therefore, one can judge the category of the whole face figure.

4 Experiment Analysis

To verify the recognition performance of the algorithm proposed in our paper, we did simulated analysis based on Yale face database. This experiment applied imresize function to resize each face image to 128×128 pixel and chose both 4×4 and 8×8 dividing algorithm. By comparing with popular classifiers SVM, SRC, and ASRC algorithm described in document [7], we can verify the performance of the algorithm.

4.1 Experiment Environment

Hardware environment: a 64-bit win10 PC with 8 GB RAM and Core i5 processor.

Software environment: MATLAB 7.0.

4.2 Experiment Results and Analysis

In this experiment, we select each category of face image with m = 3,4,5,6 from Yale face database as training samples. Considering the rationality of random sampling, repeat each algorithm for ten times by obtaining different quantity of training data. Use the average value as the final result. The identification rate of each algorithm in Yale face database is listed in Table 1, where 4×4 and 8×8 represent using 4×4 uniformly blocking method and 8×8 uniformly blocking method.

Table 1. Identification rate of each algorithm in Yale face database

Training size	Identification rate (%)				
	SVM	SRC	ASRC	4×4	8×8
3	77.46	81.13	85.92	86.83	87.67
4	82.24	85.42	86.67	89.14	90.38
5	86.67	88.46	90.19	91.22	92.78
6	88.67	90.53	91.74	93.45	94.75

According to the data in Table 1, when the training size of samples is 6, the 8×8 uniformly blocking algorithm has the highest identification rate at 97.75%. It makes a 4.22% improvement compared to traditional SRC algorithm. By analyzing the experimental data in Yale face database, one may find that as the training size increases, the two blocking algorithms performs better than other three algorithms. Therefore, the experiment verifies the accuracy of the proposed algorithm.

In the efficiency measurement, we compared three different algorithms: PCA combined with SVM, PCA combined with SRC, and uniform blocking algorithm. We divided 100 experimental data into 10 testing set, and measured the average classification time per set. The experiment result is shown in Fig. 2.

Generally, the uniform blocking algorithm has a medium classification efficiency, with average rate between 1.8 and 2.2 s. It is faster than PCA combined with SVM (between 2.4 and 2.8) and a little slower than PCA combined with SRC (between 1.4 and 1.8). As one can seen, its computation complexity did not increase too much due to its accuracy in image reconstruction.

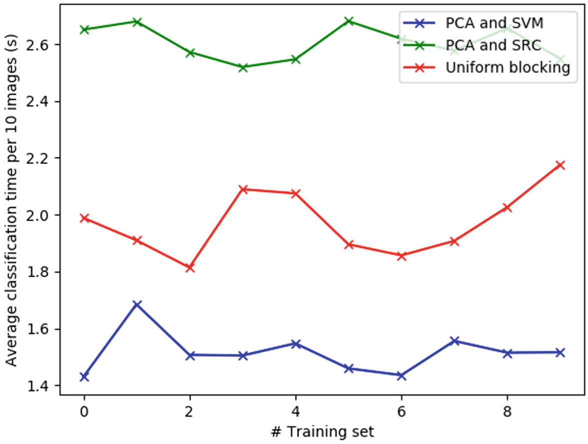

Fig. 2. Efficiency comparison between PCA combined with SVM, PCA combined with SRC, and Uniform blocking algorithm.

5 Conclusion

This paper utilizes the local sparsity within figures and the non-local similarity among image blocks to improve the performance of reconstruction and optimize the problems of solving minimal l_1-norm with respect to sparse coefficient in traditional SRC algorithm. By comparing the minimal error between testing sample figures and reconstructed sample figures, the algorithm categorizes image blocks based on majority vote, and finally realizes the recognition of face images. Experimental results show that the improved uniformly blocking SRC algorithm has an obvious improvement in face recognition rate.

References

1. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
2. Fowler, J.E., Mun, S., Tramel, E.W.: Multiscale block compressed sensing with smoothed projected landweber reconstruction. In: Signal Processing Conference, Barcelona, pp. 564–568 (2011)
3. Liu, Y., Zhang, C., Kim, J.: Disparity-compensated total-variation minimization for compressed-sensed multiview image reconstruction. In: IEEE International Conference on Acoustics, Speech and Signal Processing, South Brisbane, pp. 1458–1462 (2015)
4. Mun, S., Fowler, J.E.: Block compressed sensing of images using directional transforms. In: IEEE International Conference on Image Processing, p. 547. IEEE (2010)
5. Tramel, E.W., Fowler, J.E.: Video compressed sensing with multihypothesis. In: Snowbird: Data Compression Conference, pp. 193–202. IEEE Computer Society (2011)

6. Zhang, J., Zhao, D., Gao, W.: Group-based sparse representation for image restoration. IEEE Trans. Image Process. **23**(8), 3336–3351 (2014)
7. Wang, J., Lu, C., Wang, M., et al.: Robust face recognition via adaptive sparse representation. IEEE Trans. Cybern. **44**(12), 2368–2378 (2014)