

Low Spatial Resolution Face Recognition Based on Compressive Sensing

Xiao Hu, Shaohu Peng, Jiyong Yan, Zhen He

Department of Mechanical and Electric Engineering, Guangzhou University, Guangzhou, 510006, China

Keywords: Low spatial resolution, compressive sensing, convex optimization, principal component analysis.

Abstract. In order to effectively increase robust to recognize low spatial resolution face, this paper tried to take compressive sensing(CS). Firstly, all train face images or their corresponding feature vectors were taken to form sparse representation matrix. Secondly, test face's sparse coefficients were estimated by convex optimization. Lastly, the test face was decided as the class with minimum residuals. Two face databases (AT&T and AR) were employed to evaluate the performance of some CS algorithms such as SRC, RSC and CRC. The experiments showed that compared with PCA and FLD, the CS algorithms increased recognition rate for low resolution.

Introduction

During the past ten years, video surveillance has widespread been applied in many fields because of the increasing need for security[1][2]. Video faces have therefore more and more been employed to identify who is here. However, most video faces are low resolution (LR) faces[3] which include less discernible information across persons than high resolution (HR) faces do[4]. As a result, these traditional algorithms, which perform well on HR face recognition, become poor when they are applied in LR face recognition. Hence, it has become an important and challenge task to increase recognition rate of video face recognition for LR faces. So far, many methods have been proposed to enhance the robustness of LR face recognition. It was easy thought to recover the lost information by reconstructing HR face from LR face. A simple and fast method was interpolation such as bilinear, cubic and spline, however it was unavoidable that these recovered pixels were different from these real pixels at their corresponding position. Lee et al. [5] took support vector data description to reconstruct normal resolution face from LR face. In the proposed method, one spherical decision ball was firstly modeled from these given prototype facial images, and then each degraded LR image was projected onto the spherical decision boundary. At last, the corresponding HR facial image was synthesized. Super resolution (SR)[3]was another method which to recover loss information from a single or a set of LR face image. Many SR algorithms have been proposed so far, including the frequency domain method, the projection onto convex sets method, and the maximum posteriori (MAP) method. However, SR was an ill-posed inverse problem, in that the number of possible solutions given the available data is too great. Zhang et al.[4] proposed a morphable model space based face SR, which construct the HR information both required by reconstruction and recognition directly in the low dimensional feature space. However, SR method needs a series of LR face image.

In 2006, David L. Donoho et al. introduced[6] compressive sensing (CS), on which some signals could be measured by using less pixel than that Nyquist theorem need. After that, John Wright et al.[7] firstly took CS into face recognition, and proposed a sparse representation based classification (SRC). Later, zhang et.al[8][9]developed a collaborative representation based classification (CRC). In this paper, these CS algorithms including SRC, RSC and CRC were applied into LR face recognition.

Compressing Sensing

Sparse Representation Matrix. After expanded by an orthogonal basis, a signal $f \in \mathcal{R}^N$ could be represented as

$$f = \Psi c = \sum_{i=1}^N c_i \psi_i \quad (1)$$

Here, Ψ was a $N \times N$ orthogonal matrix and ψ_i was the i th column vector in Ψ ,

$$\Psi = [\psi_1, \psi_2, \dots, \psi_N] \quad \psi_i \in \mathcal{R}^N \quad (2)$$

$c = [c_1, c_2, \dots, c_N]$ was coefficient vector,

$$c = \Psi^{-1} f \quad (3)$$

If $0 < p < 2$ and $R > 0$, l_p -norm $\|c\|_p$, $1 \leq p \leq \infty$, satisfied the following inequality,

$$\|c\|_p = \left(\sum_{i=1}^N |c_i|^p \right)^{1/p} \leq R \quad (4)$$

When $f(t)$ has a sparse expansion, the small coefficients can be discarded without much perceptual loss. $f(t)$ was reconstructed by these K largest values in c ,

$$f_K = \Psi c_K \quad (5)$$

Here c_K was the vector of coefficients with all but the largest K set to zero. The vector was sparse in a strict sense since all but a few of its entries are zero, therefore it was called K -sparse. Though natural signal was not necessary sparse signal, its coefficients were K -sparse after the signal had translated.

Compressing Sensing. In order to include all information of one signal, under Nyquist theorem N sample points $g(t)$ need be acquired while under compressing sensing only M ($M \ll N$) sample points $f(t)$ were acquired to measure the same signal.

$$g(t) = \Phi c \in \mathcal{R}^M \quad (6)$$

$g(t)$ was called measurement vector, and $\Phi \in \mathcal{R}^{M \times N}$ was measurement matrix or projection matrix [6]. Between $g(t)$ and $f(t)$, there were translation relation.

$$g(t) = \Phi c = \Phi \Psi^{-1} f(t) \quad (7)$$

Because of $M \ll N$, $f(t) \in N$ was compressed into $g(t) \in M$, it was therefore called compressing sample.

Signal Recovery. $f(t)$ could be recovered from $g(t)$ according to Equ.(7) and Equ.(6). As the above mentioned c_K with K -sparse, one could obtain its solution according to l_0 -minimization problem, that is

$$\min \|c_K\|_0 \quad \text{subject to} \quad \Phi c_K = g \quad (8)$$

However Equ.(6) was an underdetermined linear system, lots of coefficient vectors would be gained from $g(t)$. If the solution c_K sought was sparse enough, the solution of the l_0 -minimization problem was equal to the solution to the following l_1 -minimization problem,

$$\min \|c_K\|_1 \quad \text{subject to} \quad \Phi c_K = g \quad (9)$$

Face Recognition Based on CS

Sparse Representation based Classification. John Wright[7] firstly introduced compressing sensing into face recognition to propose SRC, whose sparse representation matrix consisted of face images in train set. Supposed $\psi_{ij} \in \mathcal{R}^M$ was a column vector which was converted from the j th face image of the i th class. There were J train faces for each class face, $\Psi_i \in \mathcal{R}^{M \times J}$ therefore was representation matrix of the i th face.

$$\Psi_i = [\psi_{i,1}, \psi_{i,2}, \dots, \psi_{i,J}] \quad (10)$$

All C classes of train faces were united into one matrix

$$\Psi = [\Psi_1, \Psi_2, \dots, \Psi_i, \dots, \Psi_C] \quad (11)$$

y , one of the i th class of test faces, was expanded into

$$y = a_{i,1}\Psi_{i,1} + a_{i,2}\Psi_{i,2} + \dots + a_{i,J}\Psi_{i,J} \quad (12)$$

$a_{i,j}$ was coefficient or scalar corresponding to atom $\Psi_{i,j}$. Hence, for a test face, y , its coefficient vector to Ψ was

$$a = [0, 0, \dots, 0, a_{i,1}, a_{i,2}, \dots, a_{i,J}, 0, \dots, 0] \quad (13)$$

$a \in R^n$, $n=C \times J$, only these entries associated with the i th class might be not 0. Thus, SRC algorithm is constituted by these following steps.

Step 1: To constitute Ψ , and all column vectors in Ψ were normalized according to l_2 norm.

Step 2: To calculate sparse representation coefficient vector \hat{a} of test face y according to Equ.(9).

Step 3: To compute the residuals

$$r_i(y) = \|y - \Psi \delta_i(\hat{a})\|_2 \quad i = 1, 2, \dots, C \quad (14)$$

Step 4: Identity (y) = $\arg \min[r_i(y)]$.

Fig.1 showed a valid recognition example.

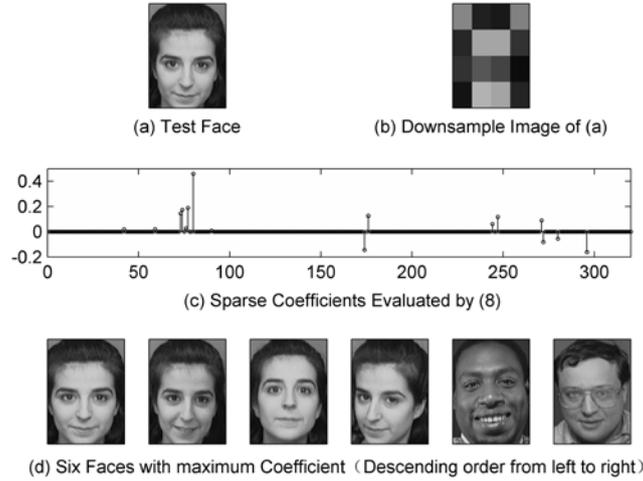


Fig.1: A Valid Face Test

Robust Sparse Coding. From maximum likelihood estimation solution of the sparse representation, whether \hat{a} which were estimated according to l_2 or l_1 norm was fidelity on its real value depended on the assumption that the coding residual follows Gaussian or Laplacian distribution. In real circumstance, that is not. Therefore, Yang et al. developed RSC[7].

$$\min_{\hat{a}} \sum_{i=1}^n \rho(y - \phi_i \hat{a}) \quad s.t. \quad \|\hat{a}\|_1 \leq \varepsilon \quad (15)$$

Here, ρ was a parameter relative to coding residual distribution.

Collaborative Representation Coding. Dictionary subsets Ψ_i and Ψ_j are not incoherent but might be highly correlated. As a result it was unstable to classify faces. Therefore, Lei Zhang et al.[8] introduced CRC by which a regularized least square method was employed,

$$(\hat{a}) = \arg \min_a \left\{ \|y - \Psi a\|_2^2 + \lambda \|a\|_2^2 \right\} \quad (16)$$

Here, λ is the regularization parameter, whose role was to make the least regularization stable and introduce a certain amount of sparsity to the solution.

Computational examples and analysis

Two publicly available face databases were taken to do experiment: one was AT&T and the other was AR. Here, principal component analysis(PCA) and fish discriminative analysis(FLD) [10] were taken to compare with CS algorithms.

AT&T Database.In AT&T database, there were ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions and facial details. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The size of each image was 112×92 pixels, with 256 grey levels per pixel. In this paper, in order to gain low quality face with low resolution, all images were converted into other five sizes by keeping the centre pixel (56,46) and down sampling from the pixel to the outside. Fig.2 showed original faces of one subject from AT&T database and their down-sampled faces.



Fig.2: Original faces of one subject from AT&T database and their down sampled face images. Spatial resolutions from top to bottom are respectively 112×92 , 56×46 , 37×31 , 23×19 , 13×11 and 11×9

In this experiment, for every subject 8 face images random extracted from 10 face images were taken to train all algorithms applied here, and other 2 face images were used to test these algorithms. All algorithms were repeatedly done 20 times. At last, the average recognition rates were listed in Table 1 which showed that CS algorithms performed better than FLD and PCA for LR faces. For these CS algorithms would become infeasible when spatial resolution were above 37×31 . Hence, only these recognition rates from low resolution faces were listed in Table 1 for CS algorithms.

Table 1: Accurate recognition rate (%)

Down sample ratio	Resolution	FLD	PCA	SRC	RSC	CRC
	112×92	93.1	90.85			
3	37×31	91.4	90			
5	23×19	90.5	89.2	89.4	89.17	84.20
7	16×13	84.3	87.8	86.06	88.20	83.90
9	13×11	62.3	77.9	81.81	82.53	79.44
11	11×9	60.7	66.5	77.5	75.92	78.71

AR Database. The AR database consisted of over 4000 frontal images for 126 individuals (70 male and 56 female). For each individual, 26 pictures were taken in two separate sessions. These images included more facial variations, including illumination change, expressions and facial disguises comparing to the AT&T database. In this paper, 118 individuals were chosen by rejecting the other subjects whose images were of faultiness. For each employed subject, 14 face images with only illumination change and expressions were selected: the seven images from Session I for training and the other seven from Session II for testing. There are four neutral faces with different lighting conditions and three faces with different expressions. These images were cropped with dimension 120×100 and converted to gray scale. For all face images, their eyes were located by

hand. And then all images were down sampled again according to eye distance (ED), and the resolutions and the corresponding ED were listed in Table 2.



Fig.3: Original faces of one subject from AR database and their down sampled face images. Spatial resolutions from top row to bottom are respectively 120×100 , 60×50 , 30×25 , 15×12 and 6×5

Table 2: EDs and their corresponding spatial resolution

ED	40	20	10	5	2
Spatial Resolution	120×100	60×50	30×25	15×12	6×5

Fig.3 showed original faces of one subject from AR database and their down-sampled faces. Spatial resolutions from top row to bottom are respectively 120×100 , 60×50 , 30×25 , 15×12 and 6×5 . In order to avoid not getting results due to high dimension face during running CS algorithms, in this experiment all CS algorithms were divided into two steps. In the first step, all face images were replaced by their feature vectors gained by PCA or FLD, and in the second step, CS algorithms were applied to these feature vectors with low dimension. Hence, there were six classes of CS algorithms by assembling CS algorithms (SRC, RSC and CRC) with PCA or FLD. However, only 4 combination methods (showed in Fig.3) can get reasonable results.

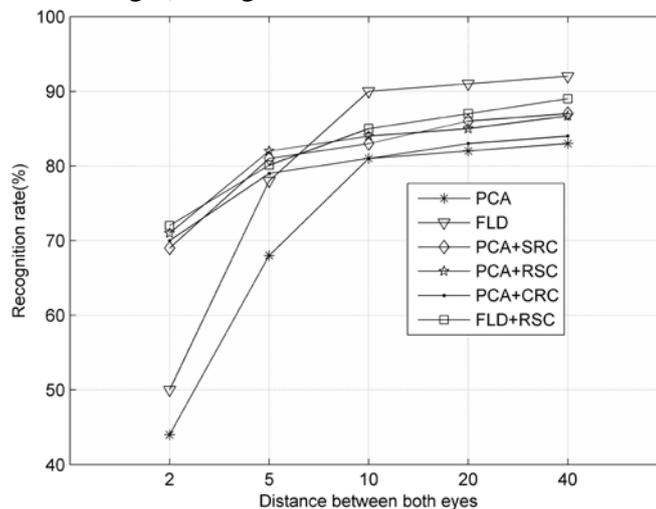


Fig.3: Results of several algorithms to recognize face from AR databases.

It was deduced from Fig.3 that when ED was above 5, FLD performed best among all above-mentioned algorithms, and when ED was below 5, the performance of FLD and PCA had become very worse, their recognition rates were below 50%. Though the recognition rates gained by CS algorithms were between those gained by FLD and PCA when the spatial resolution were above 30×25 , it was fortunately found that CS algorithms took on strong robust as the spatial resolution became lower and lower.

Conclusions

As video monitoring system was wider and wider applied, video face recognition were become more and more important. However, in real-time environment low-spatial resolution faces would occupy much ratio among video faces captured from surveillance and hindered intelligent surveillance system from being applied wide, therefore it was an important task how to increase the robust of face recognition for low spatial resolution. In this paper, all experiments told us that though it might take CS algorithms more time to recognize face, CS algorithms can increase the robust to recognize low spatial resolution face. Therefore, CS algorithms were effective.

Acknowledgements

This project was supported by Natural Science Foundation of Guangdong Province, China (No.S2013010013511), and Science and Technology Planning Project in Guangzhou, China (No.2014J4100127).

References

- [1] Hu, X., Peng, S. H., Yan, J.Y. & Zhang, N., Fast Face Detection Based on Skin Color Segmentation Using Single Chrominance Cr, *IEEE, CISP 2014*, October, Dalian, China, pp.789-794,2014.
- [2] Hu, X., Yu, W.X. & Yao, J., Face Recognition Using Binary Structure-Based Feature Selection, *Journal of Applied Sciences*, 28(3), pp.271-275, 2010.
- [3] Clinton, F., Frank, L., Vinod, C. & Sridha, S., Evaluation of image resolution and super-resolution on face recognition performance, *J. Vis. Commun. Image R.*, 23, pp.75–93, 2012.
- [4] Zhang, D., Jiazhong He, J.Z., & Du, M.H., Morphable model space based face super-resolution reconstruction and recognition, *Image and Vision Computing*,30,pp.100-108, 2012.
- [5] Lee,S.W., Park,J.& Lee,S.W., Low resolution face recognition based on support vector data description, *Pattern Recognition*, 39, pp. 1809-1812,2006.
- [6] David, L. D., Compressed Sensing, *IEEE Transaction on Information Theory*, 52(4), pp.1289-1306, 2006.
- [7] Wright, J., Yang, A.Y., Ganesh, A.& Sastry, S.S, Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31,pp.210-227,2009.
- [8] Zhang, L., Yang, M. & Feng, X. Sparse Representation or Collaborative Representation: Which Helps Face Recognition? *ICCV 2011*, pp.625-632, 2011.
- [9] Yang, M., Zhang, L., Feng, X. & Zhang, D., Sparse Representation based Fisher Discrimination Dictionary Learning for Image Classification, *International Journal of Computer Vision*, 109(3), pp. 209-232, 2014.
- [10]Hu,X., Yu W.X. & Yao, J., Multi-oriented 2DPCA for Face Recognition with One Training Face Image per Person, *Journal of Computational Information Systems*, 6(5),pp.1563-1570, 2010.