Using the idea of global-local to perform collaborative representation in face recognition

Chen Chao, Huang Jiancheng, Wang Zihao

2017, School of Engineering, Nanjing University, Nanjing 210046

Abstract: We study the problem of reliable face recognition in gray-scale images. This paper, we propose a global-local features based face recognition method. The proposed method involves three steps and it is an improvement to well-known collaborative representation method. The first step determines a linear combination of all training samples which roughly equals test sample and compute residuals of each class. The second and third step break both the training samples and test samples into fixed size blocks, again find the linear combinations of corresponding block and compute the sum of the residuals for each other. Finally, the method adds all residuals and labels the test sample with the class having least residual. The idea derived from the ensemble learning. The rationale is that it can make wiser decision on the combination information about the whole image and local feature. Our method is evaluated here on Extended Yale B and AR datasets, and is seen to perform well under varying conditions.

Key words: Information fusion; Face recognition; sparse representation; convex optimization

Introduction

As one of the most intensively investigated topics, face recognition (FR) has aroused broad interests in signal processing and computer vision. Meanwhile a great variety of face representation and classification approaches are developed. Wright *et al*.[1] reported a very conspicuous work by using sparse representation based classification (SRC), which surmises that face images of the same class map to the same face subspace. The principal idea of SRC is to find a sparse linear combination of all the training samples which can express a given test sample. the representation coefficients of training samples are supposed to be zero except those with class label same as the test sample . SRC makes a huge success in FR, and it promotes the research of face recognition based on sparsity. Zhang *et al*.[2] presented Correlation Representation based Classifier (CRC) in consideration that the L2-norm-based sparsity constraint plays a greater role than the traditional L1-norm-based constraint.

Motivated by the idea of ensemble learning and global-local relationship, we present local-whole collaborative representation classifier (GLCRC). Traditional sparse representation methods focus more on the holistic characteristic of the face. However, recently more attempts are made to exploit the advantage of local features, which are considered more robust in certain cases. Therefore, It is wiser to combine both the local and global features smartly in face recognition system. Our experiments have shown that different noises in the image samples cause different residuals. The global part shows stronger robustness to impulse noise, while the local parts are more robust to the variations of occlusion and illumination.

As shown in our experiments, the resulting algorithm works well in complex scenes. The rest part of this paper is organized as follows: Section 2 elaborates on our method. In section 3, we present the experimental results. Finally, section 5 offers our conclusion.

1 The global-local collaborative representation method

This section describes the proposed global-local collaborative representation method in detail. Before proceeding, it is imperative to define some terms for use throughout the remainder of this paper. Assume that there are n training samples, x_1, x_2, ... x_n. Each sample is supposed to a one-dimensional column vector. Hence the training set is a matrix composed of n column vectors. All samples fall into C classes.

1.1 The first stage of GLCRC

In the first stage of GLCRC, collaborative representation classifier is employed on the overall samples. The stage supposes that the following equation is satisfied:

$$\mathbf{y} \approx a_1\mathbf{x_1} + \cdots + a_n\mathbf{x_n} \tag{1}$$

where y is the given test sample and a_i (i=1, 2, ..., n) are the coefficients of x_i. We rewrite Eq.1 as

$$\mathbf{y} \approx \mathbf{X}\alpha \tag{2}$$

where $\mathbf{X} = [\mathbf{x_1}, ..., \mathbf{x_n}]$ and $\alpha = [a_1, ..., a_n]$. With the hope that the coefficent vector α is sparse, the problem one would like to sovle can be stated as

$$\min \|\alpha\|_1 \quad s.t. \|\mathbf{y} - \mathbf{X}\alpha\|_2 \leq \varepsilon. \tag{3}$$

where ε is a small number. To make the problem easier to solve and computationally more efficient, ℓ_1 is replaced by ℓ_2:

$$\min \|\alpha\|_2 \quad s.t. \|\mathbf{y} - \mathbf{X}\alpha\|_2 \leq \varepsilon. \tag{4}$$

Eq.4 can be rewritten as

$$\min \lambda \|\alpha\|_2^2 + \|\mathbf{y} - \mathbf{X}\alpha\|_2^2 \tag{5}$$

where λ is a positive constant. We obtain the solution of α using

$$\widetilde{\alpha} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \tag{6}$$

Finally, we compute the residual of each class using

$$e_r = \frac{\|\mathbf{y} - \mathbf{X}\sigma_r(\widetilde{\alpha})\|_2}{\|\sigma_r(\widetilde{\alpha})\|_2} \tag{7}$$

where $\sigma_r(\widetilde{\alpha})$ denotes the coefficients of rth class and e_r stands for the residual of rth class The first stage considered the correlation between the whole face. Therefore, it is robust to pixel noise to some extent.

1.2 The second stage of GLCRC

The second stage divides both the training samples and test sample into K^2 nonoverlapping regions with the same size. We set $K = 2$, thus each region is a rectangle of height $\frac{H}{K}$ and width $\frac{W}{K}$, where H and W are the height and width of original image. Then we regard each region of the query sample as a local query sample $\mathbf{y}_k$ and the collection of corresponding region of all the training samples as a local dictionary $\mathbf{X}_k$. Just like the first stage, we calculate the representation coefficient with respect to the local dictionary for each local query sample using

$$\widetilde{\alpha}_k = (\mathbf{X}_k^T \mathbf{X}_k + \lambda \mathbf{I})^{-1} \mathbf{X}_k^T \mathbf{y}_k \quad k = 1, 2, \cdots, K^2 \tag{8}$$

where $\mathbf{X}_k$ stands for the collection of the kth region of each training sample and $\mathbf{y}_k$ denotes the kth region of a query sample.

Then we can compute the residuals for each region separetely using

$$\sigma_{kr} = \frac{\|\mathbf{y}_k - \mathbf{X}_{kr} \widetilde{\alpha}_{kr}\|_2}{\|\widetilde{\alpha}_{kr}\|_2} \quad k = 1, 2, \cdots, K^2 \tag{9}$$

where σ_{kr} stands for the residual of rth class for the kth region. And then we compute their mean value in the second stage like:

$$\sigma_r = \frac{\sum_k \sigma_{kr}}{K^2} \quad k = 1, 2, \cdots, K^2 \tag{10}$$

The second stage integrates information between parts of the image and it is less likely to be effected by occlusion.

1.3 The third stage of GLCRC

The third stage is similar to the second stage, but it divides image into $\hat{K}^2$ instead, where $\hat{K}$ is greater than K. We set $\hat{K} = 4$. The same as the second stage, this stage compute the residual of each class by:

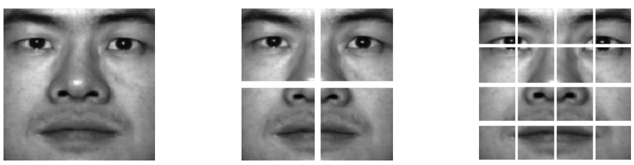
$$\widetilde{\alpha}_{\hat{k}} = (\mathbf{X}_{\hat{k}}^T \mathbf{X}_{\hat{k}} + \lambda \mathbf{I})^{-1} \mathbf{X}_{\hat{k}}^T \mathbf{y}_{\hat{k}} \quad \hat{k} = 1, 2, \cdots, \hat{K}^2 \tag{11}$$

$$\xi_{r\hat{k}} = \frac{\|\mathbf{y}_{\hat{k}} - \mathbf{X}_{\hat{k}r} \widetilde{\alpha}_{\hat{k}r}\|_2}{\|\widetilde{\alpha}_{\hat{k}r}\|_2} \quad \hat{k} = 1, 2, \cdots, \hat{K}^2 \tag{12}$$

$$\xi_r = \frac{\sum_{\hat{k}} \xi_{r\hat{k}}}{\hat{K}^2} \quad \hat{k} = 1, 2, \cdots, \hat{K}^2 \tag{13}$$

The third stage mines the local information more fully. It performs better in occlusion issue. However, it does not mean the greater the $\hat{K}$, the better. Too fine a division may break the property of face subspace. In the extreme situation where each block contains only one pixel, it turns into the nearest neighbour approach (NN), i.e. compute the corresponding pixel distance and determine the class with the least total distance.

1.4 Summary of GLCRC



(a) (b) (c)

Figure 1: Image segmentation method in three stages. (a) 1×1 in firts stage. (b) 2×2 in second stage. (c)4×4 in third stage.

Fig.1 shows the way we divide the face image in three stages respectively. The proposed algorithm is summarized in the following steps:

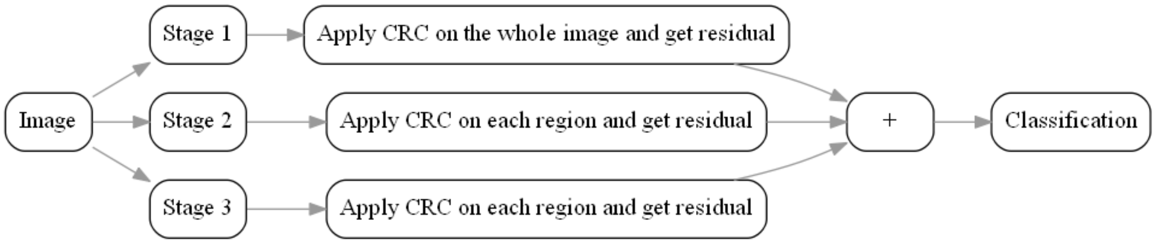


Figure 2: Illustration of GLCRC process

1. Preprocessing: Convert both the training and testing samples into grayscale ones and normalize them.
2. Stage 1: Apply CRC on the whole image. get the residual of each class.
3. Stage 2: divide the images into K^2 regions. Apply CRC on each region of image. get the residual of each class.
4. Stage 3: divide the image into $\hat{K}^2$ regions. get the residual of each class.
5. Classification: add the residuals of each class in the three stages up and assign the test sample into the class with the least residual.

$$s_r(\mathbf{y}) = e_r + \sigma_r + \xi_r \tag{14}$$

$$Label(\mathbf{y}) = \arg\min_r s_r(\mathbf{y}) \tag{15}$$

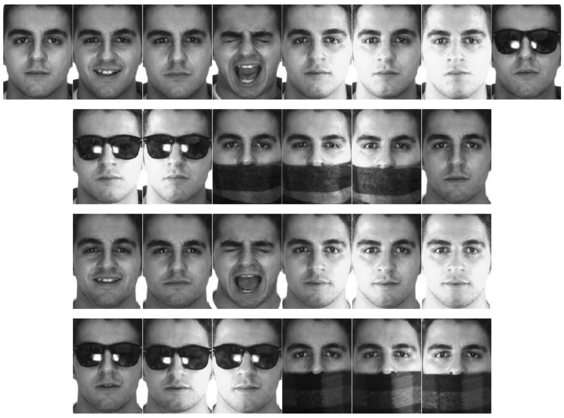
Fig.2 illustrates the process of GLCRC.

2 Experiment Results

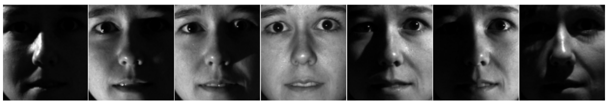
We set up a series of experiments on the AR[3] and Extended Yale B[4] face dataset. Fig.3 shows some face images from Extended Yale B dataset and AR dataset. Besides GLCRC, another three methods including CRC, BCRC2, BCRC4 are employed for comparison. CRC is actually the method used in the first stage of GLCRC. BCRC2 and BCRC4 stands for the method used in the second and third stage respectively. For consistency, we set the parameters $\lambda = 1$. All images used are resized to 100×100. The experiments are implemented in python2.7 on computer with 3.60 GHz CPU and 8.00 GB RAM.

2.1 FR With Illumination Changes

We perform experiments on Extended Yale B to validate the robustness of our method to illumination changes. The whole dataset which consists of 2414 frontal face images over 38 individuals, is divided into five subsets in accordance with the illumination conditions in images[4]. From subset 1 to 5, the face images characterize slight-moderate-severe illumination changes. The images from subset 1 (about 7 images per person) are used for training, while the rest are for testing. Fig.4(a) shows the recognition rates of four methods using various number of training samples. It can be clearly seen that BCRC4, GLCRC, BCRC2 obtains higher recognition rates comparing to CRC. The main reason for encouraging performance of methods based on local feature can be attributed to the fact that some sub-block are little polluted by uneven illumination hence maintain subspace property. Taking advantage of local-feature based method, GLCRC can achieve competitive results under the circumstance of illumination changes.

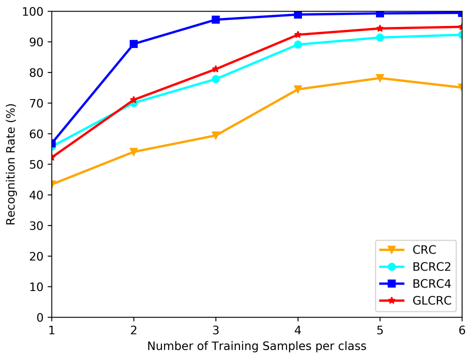


(a)

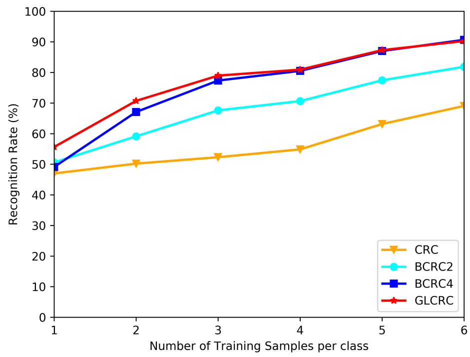


(b)

Figure 3: Some test images used in our experiments. (a) face images of one individual from AR database. (b) face images from Extended Yale B of one individual under seven kinds of light conditions.



(a)



(b)

Figure 4: The recognition rates(%) of different methods on (a) Extended Yale B database, (b) AR database.

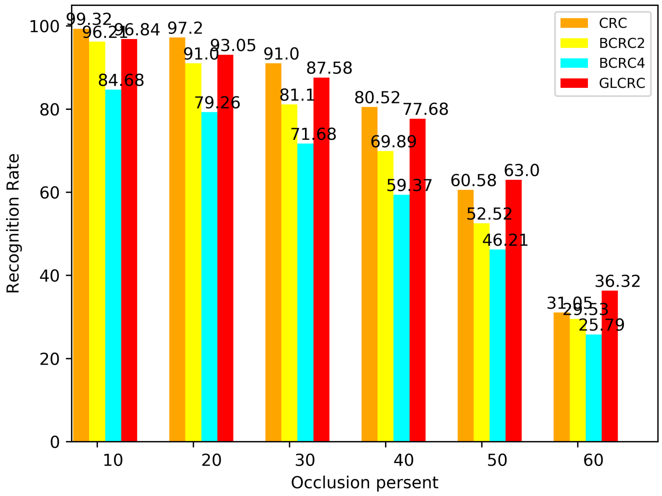
2.2 FR With Contiguous Occlusion

We first conducted experiments on AR database to verify the robustness of GLCRC to various contiguous occlusion. The face images of 100 individuals are used. We select 1 to 6 clean face images per person as training samples, while the occluded images remained as testing samples. Fig.4(b) exhibits the experiments results of all methods. We can find that GLCRC achieves best performance.

In real life, besides sunglasses and scarf, occlusion may be other stuff such as hair, mask, leaves, etc. It is hard to predict the location and size of occlusion. Therefore, it is necessary to verify the robustness of face recognition method with various occlusion. In the second experiment, we design a random block occlusion on subset 1 of Extended Yale B database. the first image of each class is used for training and the rest for testing. We set increasing levels of square block occlusion, from 10% to 60%, using an unrelated baboon image which is employed in many studies [5, 6, 7]. Fig.5(a) shows the sample image with various levels of random occlusion. The results of experiment are shown in Fig.5(b). We can observe that global-feature based method obtain competitive performance. Specially, GLCRC is more stable with the increase of occlusion area. It outperforms other methods when most area of image are blocking. This mainly owing to the superiority of integrated information about global and local features.

All These results demonstrate that GLCRC is powerful and robust to complex facial occlusion.

(a)



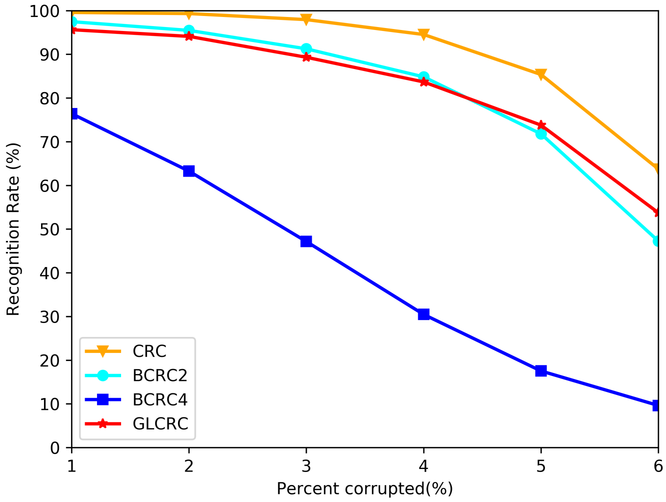
(b)

Figure 5: (a)the baboon image for occlusion and face images from Extended Yale B with random occlusion (from 10% $\sim$ 60%). (b) The recognition rates(%) of different methods on Extended Yale B with various levels of random occlusion.

2.3 FR With Pixel Corruption

Pixel corruption exists in image collection or transmission when environmental noise affects. In this experiment, we investigate the performance of GLCRC for pixel noise. Different levels of random noise are imposed on subset 1 of Extended Yale B database. We choose one image per person for training and the rest for testing. Fig.6(a) shows the sample image with 10% $-$ 60% pixels contaminated. The recognition rates of four methods are shown in Fig.6(b). Intuitively, CRC and GLCRC seem no so sensitive to pixel corruption. With 60% pixels corrupted, CRC obtains surprising performance with 64% and GLCRC ranks the second with 54%. It is may be that the global structure of face is still preserved even when most pixel are corrupted. The recognition rate of BCRC4 drops rapidly with the increase of corruption percentage. Though GLCRC is slightly affected by the part of BCRC4, it still does well among these four methods. These experiment results prove the robustness of our method to pixel corruption.

(a)



(b)

Figure 6: face images from Extended Yale B with seven levels of pixel corruption (i.e, 0% $\sim$ 60%). (b) The recognition rates(%) of different methods on Extended Yale B with various levels of pixel corruption.

2.4 FR With Mixed Noises

In this section, we explore the performance of our method for mixed noise. Both pixel corruption and contiguous occlusion are used on subset 1 of Extended Yale B database. Same with before experiments, the first image of each person is selected to form the training set. Besides 5% pixel corruption, We put a wide black rectangle on mouth as a scarf and a slim black rectangle on eyes as sunglasses respectively. With the purpose of comparison, the experiment with occlusion and no pixel corruption is set as well. Fig.7 shows an example face under four settings. Table1 lists the recognition rates of four methods. As we can see, in most cases GLCRC outperforms other methods except the face images with scarf. The results further confirm that GLCRC is an effective and robust FR method.

Interestingly, BCRC4 performs better on face images with the scarf than those with sunglasses, even though the scarf covers more area than sunglasses. Combined with the result of previous experiments, there is a conclusion that local-feature based sparse representation can achieve competitive performance in facial occlusion problem only if the face images are properly divided. This requires more prior knowledge about potential occlusion and add complexity to face recognition system.

3 Conclusion

In this paper, a global-local collaborative representation classifier (GLCRC) is proposed for robust face recognition. The method used three stages to classify face images. Its first stage performs collaborative representation classification on the whole image, while the second and the third perform collaborative representation classification on divided subimages. The approach is simple, theoretically reasonable, and more importantly, easily programmed. Experiments under different settings show the

(a) (b) (c) (d)

Figure 7: face images from Extended Yale B with under different experiment settings. (a) with mouth obscured. (b) with mouth obscured and 5% pixel corruption. (c) with eyes obscured. (d) with eyes obscured and 5% pixel corruption.

Method	scarf	scarf+5% pxiel noise	sunglasses	sunglasses+5% pxiel noise
CRC	39.47	41.53	77.89	77.58
BCRC2	75.26	74.84	79.47	78.73
BCRC4	**83.68**	67.52	78.95	65.47
GLCRC	78.95	75.00	**83.16**	**81.32**

Table 1: The recognition rates(%) of CRC, BCRC2, BCRC4, GLCRC under different experiment settings.

promising accuracy the proposed method obtains in dealing with complex noise.

In the future, we plan to employ sparse representation classifier on deep facial features which are extracted by some pre-trained convolutional neural networks.

4 Acknowledgement

This paper is supported particially by $\cdot$.

References

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. IEEE PAMI, 31(2):210–227, 2009.

[2] R. Rigamonti, M. Brown, and V. Lepetit, "Are sparse representations really relevant for image classification?" in IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1545–1552.

[3] A. M. Martinez and R. Benavente, "The AR face database," CVC, Barcelona, Spain, Tech. Rep., 1998.

[4] 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 Trans. Pattern Anal. Mach. Intell, no. 6, pp.643–660, 2001.

[5] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," IEEE Trans. Image Process., vol. 22, no. 5, pp. 1753–1766, 2013.

[6] J. Chen, J. Yang, L. Luo, J. Qian, and W. Xu, "Matrix variate distribution-induced sparse representation for robust image classification," IEEE Trans. Neural Netw. Learn. Syst., vol. 26, no. 10, pp. 2291–2300, 2015.

[7] J. Zheng, K. Lou, X. Yang, C. Bai, and J. Tang, "Weighted mixednorm regularized regression for robust face identification," IEEE Trans. Neural Netw. Learn. Syst., 2019.

基于整体局部关系的协同表示人脸分类器

陈超, 黄健成, 王子豪

南京大学工程管理学院 2017 级，南京 210046

摘要　基于著名的协同表示分类器，本文提出了一种整体-局部人脸识别方法。该方法分为三步：第一步找到训练样本的一个线性组合，使它可以近似表示整个测试样本，然后计算每个类的残差。第二步和第三步将训练样本和测试样本分解成相同个数固定大小的块，对测试样本的每个块找到训练样本相应块的最佳线性组合，并计算残差之和。最后，该方法将所有残差相加，并将测试样本分配到残差最小的类中。这种思想来源于集成学习。其基本原理是将图像的整体信息与局部特征信息相结合，做出更明智的决策。在 AR 和 Extended Yale B 数据库上的实验结果表明，该方法在不同的条件下均具有较高的识别率和较强的鲁棒性。

关键词　信息融合；人脸识别；协同表示；凸优化