Blockwise Collaborative Representation based Classification via L_2-norm of Query Data for Accurate Face Recognition

Jin Hee Na and Hyung Jin Chang *

In this paper we present a new blockwise collaborative representation based classification (BCRC) with L_2-norm of test data for accurate face recognition. For training we divide images into several blocks and estimate representation coefficients of each block via L_2-norm minimization. For testing the L_2-norm of test image blocks are scaled by the trained representation coefficients. A novel classification scheme based on the L_2-norm of test blocks is proposed and this scheme is jointly applied with conventional reconstruction error based classification. Experimental results show that our proposed methods outperform other representation based methods for face recognition.

Introduction: Collaborative representation based classification (CRC) [1] was proposed by raising a question on the role of sparsity in classification. Shi *et al.* [2] have experimentally shown that enforcing sparsity onto the representation coefficients does not guarantee robustness to noise in recognizing faces. They have shown that orthonormal L_2-norm based on a standard least square approach outperforms sparse representation based classification (SRC) [3] for the face recognition problem. Furthermore, Zhang *et al.* [1] have shown that representing test data by samples from all training data makes SRC more effective. In this sense, they proposed collaborative representation based classification with regularized least square (CRC_RLS) and showed that CRC outperforms SRC in respect to accuracy and efficiency. Especially, Zhang *et al.* analyzed the working mechanisms of SRC and CRC, and showed that collaborative representation even in SRC is important for classification. Since these CRC approaches are based on L_2-norm minimization, it can have a closed-form solution which is much efficient than the time consuming L1-norm solution of SRC.

Most of previous improvements [4, 5, 6] are more focused on the modification of representation learning to increase the discrimination power while the classification schemes of those methods are still based on class specific reconstruction error. Timofte and Gool [7] suggested the different classification scheme which directly uses the absolute value of representation coefficients. However, they mentioned the scheme only for SRC and did not consider it for blockwise representation learning.

In this paper, we suggest two new classification schemes with blockwise L_2-norm of test data. Blockwise classification is proposed with the L_2-norm of test data instead of reconstruction error, and a linear combination of the blockwise classifiers construct a final global classifier. For this purpose, we divide images into several blocks and estimate the representation coefficients of blocks via L_2-norm minimization. These blockwise representation coefficients are used for classwise scaling the L_2-norm of test blocks. Then, the weighted L_2-norm of test blocks are linearly combined and the class label of test data is determined by the maximum summation of L_2-norm of test blocks. We call this proposed method as block collaborative representation based classification with L_2-norm of test data (BCRC_L_2). Additionally, we propose BCRC_L_2R which combines BCRC_L_2 and original CRC_RLS by a simple rule. Experimental results on face recognition show that BCRC_L_2 and BCRC_L_2R both outperform previous methods.

Collaborative Representation based Classification (CRC): Suppose that $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \in \mathbb{R}^{d\times n}$ is given where $\mathbf{x}_i$ is a d-dimensional training data and n is the number of data. Each $\mathbf{x}_i$ belongs to one of l classes. Let $\mathbf{X}_i \in \mathbb{R}^{d\times n_i}$ is a set of training data from class i where n_i be the number of data belonging to the class $i = 1, 2, ..., l$. Once a test data $\mathbf{y} \in \mathbb{R}^d$ comes, we code it as $\mathbf{y} = \mathbf{X}\rho$, where $\rho \in \mathbb{R}^n$ is a set of representation coefficients. The optimal solution of ρ with respect to the smallest reconstruction error is obtained by solving

$$\rho^* = \arg\min_{\rho} \|\mathbf{y} - \mathbf{X}\rho\|_2^2, \tag{1}$$

*Corresponding Author. Tel.:+44 (0)20 75946274. Address: Imperial College, South Kensington Campus, London SW7 2AZ, UK.

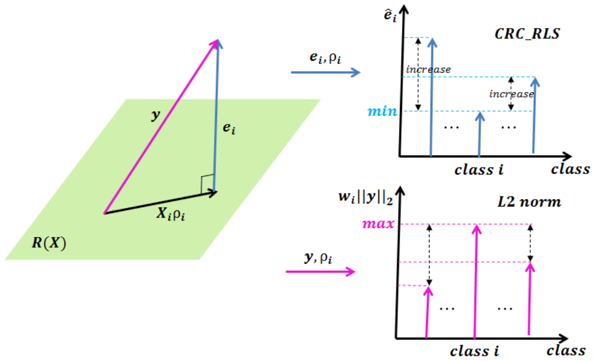

Fig. 1 *Geometric illustration of classification scheme. Left: Geometric characterization of the classification.* $\mathbf{R}(X)$ *denotes a space spanned by* $\mathbf{X}$. $\mathbf{X}_i\rho_i$ *represents a linear combination of the data in class i, and* $\mathbf{e}_i$ *is* $\mathbf{e}_i = \|\mathbf{y} - \mathbf{X}_i\rho_i\|_2^2$. *Top right: reconstruction error based classification (CRC_RLS). Bottom right:* L_2*-norm of test data based classification (proposed)*

where all training data are linearly combined to minimize least square reconstruction error of test data. It is known that the use of all training data to collaboratively represent test data alleviate the small sample size problem in face recognition [1]. The collaborative representation based classification with regularized least square (CRC_RLS) is presented in [1] with the following formulation:

$$\rho^* = \arg\min_{\rho}\{\|\mathbf{y} - \mathbf{X}\rho\|_2^2 + \lambda\|\rho\|_2^2\} \tag{2}$$

where λ is the regularization parameter. The solution in Eq. (2) can be analytically derived as

$$\rho = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda\cdot\mathbf{I})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y} = \mathbf{P}\mathbf{y}, \tag{3}$$

where $\mathbf{P} = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda\cdot\mathbf{I})^{-1}\mathbf{X}^\mathsf{T} \in \mathbb{R}^{n\times d}$ is a linear transformation matrix which can be pre-computed without considering $\mathbf{y}$. In CRC_RLS, the class label of $\mathbf{y}$ is determined by

$$class(\mathbf{y}) = \arg\min_i \|\mathbf{y} - \mathbf{X}_i\rho_i\|_2/\|\rho_i\|_2. \tag{4}$$

where ρ_i is a set of representation coefficients which length is n_i. According to a geometrical analysis in [1], the least square term in Eq. (4) is effective and robust because the class label of $\mathbf{y}$ is determined from the minimum error among classwise reconstruction errors. In addition, they mentioned that dividing the reconstruction error by $\|\rho_i\|_2$ brings additional discrimination information. Please refer to [1] for more details.

Proposed Methods: We focus on the effect of the class specific representation coefficients in Eq. (4), and propose a new classification measure based on the L_2-norm of test data rather than reconstruction error. In our work, the representation coefficients are used for classwise scaling of the L_2-norm of test data. Figure 1 shows a geometric illustration of the difference between the reconstruction error based classification and the proposed L_2-norm of test data based classification. The subfigure in the top right shows the geometrical illustration of the reconstruction error based classification, especially CRC_RLS. While class specific reconstruction error $\mathbf{e}_i$ is used as a general measurement, the use of $\hat{\mathbf{e}}_i$ in CRC_RLS increases the difference between the class specific reconstruction errors as the ρ_i corresponding to the class with minimum reconstruction error is less sparse and has larger L_2-norm than the other $\rho_j (j \neq i)$. (Namely, the dividing term by $\|\rho_i\|_2$ increases the difference between classwise reconstruction errors.) The subfigure in the bottom right shows the strategy of the proposed classification scheme. The L_2-norm of test data $\mathbf{y}$ is scaled by the weight w_i which is defined by $w_i = \|\rho_i\|_2/\|\rho\|_2$. In this scheme, the class label of test data is determined by

$$class(\mathbf{y}) = \arg\max_i w_i\|\mathbf{y}\|_2, \tag{5}$$

where class specific weighted L_2-norms of test data are compared each other and the class label corresponding to maximum L_2-norm preservation is considered as the class label of test data. It is noted that training data do not require for the classification of test data. Original

CRC_RLS requires to save both $\mathbf{P}$ and $\mathbf{X}$ for the classification of $\mathbf{y}$. Meanwhile, in our method, only $\mathbf{P}$ is required to keep from a training phase and $\mathbf{w} = \{w_i\}_{i=1}^{l}$ can be calculated fast in a test phase. Namely, Eq. (4) requires $2 \times (d \times n)$ dimensional memory on training results while the proposed classification scheme of Eq. (5) requires only $d \times n$ dimensional memory.

Based on this classification scheme, we propose two block collaborative representation based classification methods with L_2-norm: BCRC_L_2 and BCRC_L_2R. Firstly, suppose the images are divided into n_B blocks and the blockwise representation coefficients of j^{th} block $(j = 1, \cdots, n_B)$ are estimated via L_2-norm minimization:

$$\alpha_j^* = \arg\min_{\alpha_j}\{\|\mathbf{B}_j^{(\mathbf{y})} - \mathbf{B}_j^{(\mathbf{X})}\alpha_j\|_2^2 + \lambda_B\|\alpha_j\|_2^2\}, \quad (6)$$

where $\mathbf{B}_j^{(\mathbf{y})}$ is j^{th} image block of $\mathbf{y}$ and $\mathbf{B}_j^{(\mathbf{X})}$ is a set of j^{th} image blocks of $\mathbf{X}$, respectively. α_j indicates a set of representation coefficients corresponding to image block j and λ_B is the blockwise regularization parameter. We can formulate the classification scheme of BCRC_L_2 as

$$class(\mathbf{y}) = \arg\max_i \sum_{j \in \mathbf{B}^{(\mathbf{y})}} (w_{i,j}\|\mathbf{B}_j^{(\mathbf{y})}\|_2), \quad (7)$$

where $w_{i,j}$ is the weight defined as $w_{i,j} = \|\alpha_{i,j}\|_2/\|\alpha_j\|_2$ where $\alpha_{i,j}$ is a subset of α_j corresponding to the class i.

Furthermore, it is natural to consider block collaborative representation based classification with joint L_2-norm of test data and reconstruction error as follows:

$$class(\mathbf{y}) = \arg\max_i \sum_{j \in \mathbf{B}} (\gamma(w_{i,j}\|\mathbf{B}_j^{(\mathbf{y})}\|_2) - e_{i,j}(\mathbf{y})), \quad (8)$$

where γ is the balancing parameter among two measures and $e_{i,j}(\mathbf{y}) = \|\mathbf{B}_j^{(\mathbf{y})} - \mathbf{B}_j^{(\mathbf{X}_i)}\alpha_{i,j}\|_2/\|\alpha_{i,j}\|_2$, respectively. $\mathbf{B}_j^{(\mathbf{X}_i)}$ indicates a set of j^{th} block of $\mathbf{X}_i$. We call this classification scheme as BCRC_L_2R.

Experimental results: We evaluated the performance of the proposed methods using two well-known face databases: AR face database and extended Yale B face database. The AR face database consists of over $4,000$ frontal face images for 126 individuals and the extended Yale B face database consists of $2,414$ frontal face image of 38 individuals. For our experiments, we used the subset of AR face database from [1] and prepared the Yale database by using the code provided by [2]. Both database consist of gray-scale images and were resized to 60×43.

We compared the face recognition results of our method with CRC_RLS [1], relaxed collaborative representation (RCR) [4], Fisher discrimination dictionary learning (FDDL) [6] and weighted collaborative representation classifier (WCRC) [5]. RCR exploits the similarity and distinctiveness of features and multi-block RCR which uses blockwise pixel intensity as the feature is selected for our evaluation. FDDL is a variant of SRC which uses discriminative information in both sparse representation coefficient learning and identification scheme. In [5], weighted collaborative representation classifier (WCRC) has been proposed, and many variants of WCRC such as adaptive WCRC (AWCRC), kernel CRC (KCRC), kernel weighted CRC (KWCRC) and kernel adaptive WCRC (KACRC) have also been proposed. We evaluated the accuracy of WCRCs by varying the regularization parameters ($10^{-8} \sim 10^{2}$) and optimal value on each WCRC was applied. Then, best results among the WCRC variants on two databases were selected in our comparison. Except WCRC, the parameters of other methods were set as the default values suggested by authors. Eigenface with 300 dimensionality were prepared for SRC, CRC_RLS, FDDL and WCRC while 60×43 dimensional gray images were used as input of RCR and our methods because these were based on blockwise processing. For RCR and our methods, we varied the number of blocks and the highest values were selected for comparison. In addition, the robustness of recognition rates on varying the number of blocks were compared between our methods and multi-block RCR. There was no overlap between the image blocks as in [4].

Table 1 shows the accuracy comparison on the AR face database and the extended Yale B face database. For the case of AR face database, BCRC_L_2 achieved the best accuracy and BCLR_L_2R was the second best on the database. For the case of extended Yale B face database, BCRC_L_2R achieved the best accuracy and one of WCRCs achieved the second best. The recognition rate of BCRC_L_2 was slightly lower than WCRC and same with RCR. It is notable that our methods achieved the best results with only simple modifications of identification scheme while most of previous methods were based on time-consuming iterative learning to improve the accuracy of SRC or CRC. Especially, our methods are based on blockwise processing like multi-block RCR, but achieves the better recognition rates than those of multi-block RCR. In addition, for WCRCs, we selected best results among the variants of WCRC for comparison. WCRC and WCRC achieved best results on AR face database while Kernel CRC (KCRC) showed best result for extended Yale B face database.

Table 1: Face recognition accuracy

Methods	Accuracy	
	AR face database	Extended Yale B face database
CRC_RLS [1]	93.85%	97.91%
multi-block RCR [4]	96.71%	98.55%
FDDL [6]	93.56%	98.16%
WCRCs [5]	93.99%	98.82%
BCRC_L_2	**98.43**%	98.55%
BCRC_L_2R	97.71%	**100.00**%

Table 2: Recognition accuracy with varying the number of blocks

Methods	AR (Ave. ± Std.)	Extended Yale (Ave. ± Std.)
multi-block RCR [4]	94.38 ± 1.61%	95.29 ± 1.58%
BCRC_L_2	95.21 ± 3.41 %	97.63 ± 0.57%
BCRC_L_2R	**95.29** ± 2.10 %	**99.62** ± 0.26%

The recognition rates of our methods and multi-block RKR were dependent on the number of blocks. Since it was difficult to determine the optimal number of blocks, we compared the average recognition rates on the different number of blocks. In this experiments, the number of blocks, n_B, were varying from 2 (1×2, 2×1) to 64 (8×8). Table 2 shows the average recognition rates and the standard deviations. For AR face database, the standard deviation of multi-block RCR was less than our methods but the average recognition rates of ours were higher than that of multi-block RCR. For the extended Yale B face database, our methods outperformed multi-block RCR.

Conclusion: In this paper, we first proposed a blockwise CRC scheme based on the weighted L_2-norm of test data. In this BCRC_L_2, the coefficients of collaborative representation are used for determining class specific weights and scaling L_2-norm of test data. The use of L_2-norm of test data is memory efficient because it does not require training data for evaluation. Experimental results showed that the use of weighted L_2-norm of test data is able to achieve higher recognition rates on benchmark face databases than previous CRC and its variants. We additionally proposed another identification scheme (BCRC_L_2R) based on joint L_2-norm of test data and reconstruction error. As expected, this modification also achieved comparable recognition rates on two well-known face databases.

Jin Hee Na (*Samsung Electronics Co., Ltd., Republic of Korea*) and Hyung Jin Chang (*Imperial College London, United Kingdom*)

E-mail: hj.chang@imperial.ac.uk

References

1 Zhang, L., Yang, M., and Feng, X.: ‘Sparse Representation or Collaborative Representation: Which Helps Face Recognition?’, *in Proc. Int. Conf. Comp. Vis.,*, 2011.

2 Shi, Q., Eriksson, A., van den Hengel, A., and Shen, C.: ‘Is face recognition really a Compressive Sensing problem?’, *in Proc. Int. Conf. Computer Vision and Pattern Recog.*, 2011.

3 Wright, J., Yang, A. Y., Sastry, S. S., and Ma, Y.: ‘Robust face recognition via sparse representation’, IEEE Trans. Pattern Anal. Mach. Intell., 2009, **31**, pp. 210-227

4 Yang, M., Zhang, L., Zhang, D., and Wang, S.: ‘Relaxed Collaborative Representation for Pattern Classification’, *in Proc. Conf. Computer Vision and Pattern Recog.*, 2012.

5 Timofte, R., and Gool, L. V.: ‘Adaptive and Weighted Collaborative Representations for Image Classification’, Pattern Recogn. Lett., 2014, **43**, pp. 127-âĂŞ135

6 Yang, M., Zhang, L., Feng, X., Zhang, D.: ‘Sparse Representation based Fisher Discrimination Dictionary Learning for Image Classification’, Int. J. Comput. Vision, 2014.

7 Timofte, R. and Gool, L. V.: ‘Sparse Representation Based Projections’, *in Proc. British Machine Vision Conference*, 2011,