SCIENCE CHINA
Information Sciences

• HIGHLIGHT •

April 2017, Vol. 60 048101:1–048101:3
doi: 10.1007/s11432-015-0799-4

Collaborative representation Bayesian face recognition

Jufu FENG, Xiao MA* & Wenjing ZHUANG

School of Electronics Engineering and Computer Science, Peking University, Beijing 100871*, China*

Received May 19, 2016; accepted July 8, 2016; published online December 20, 2016

Citation Feng J F, Ma X, Zhuang W J. Collaborative representation Bayesian face recognition. Sci China Inf Sci, 2017, 60(4): 048101, doi: 10.1007/s11432-015-0799-4

In recent years, a series of subspace methods, named the collaborative representation based methods, have aroused researchers' interests. Inspired by the idea of sparse coding, Wright et al. [1] proposed the sparse representation based classification (SRC) method. SRC encodes a query sample as a linear combination of all subjects' training samples with sparsity constraint and then classifies it by evaluating which class has the minimal coding residual. This mechanism is named Collaborative Representation. When the training dictionary is overcomplete, SRC can achieve state-of-the-art recognition accuracy and is proven robust to occlusion and corruption. Based on SRC, Zhang et al. [2] emphasized the collaborative representation mechanism's importance and replaced the sparsity constraint by other loose constraints in SRC. A more general method, named collaborative representation based classification (CRC), was then proposed, which achieves similar recognition accuracy as SRC with much higher computational efficiency. CRC consists of two major parts: the collaborative representation model and the residual model. The collaborative representation model is defined in

$$\hat{\alpha} = \arg\min_{\alpha} \|y - X\alpha\|_2^2 + \lambda\|\alpha\|_p, \qquad (1)$$

where, y is the query sample, $X = [X_1, X_2, \ldots, X_C]$ is the training dictionary set of C subjects, and $\alpha = [\alpha_1^{\mathrm{T}}, \alpha_2^{\mathrm{T}}, \ldots, \alpha_C^{\mathrm{T}}]^{\mathrm{T}}$ is the corresponding representation coefficients. When $p = 1$, CRC turns to SRC [1]. When $p = 2$, CRC turns to collaborative representation based classification with regularized least square (CRC-RLS) [2].

Collaborative representation is a competitive mechanism. Samples of all classes contribute competitively to represent y. If one class contributes more, the other classes contribute less. With sparse constraint, lower reconstruction error will be achieved with less elements by the class y truly belongs to. Moreover, the collaborative representation can be viewed as a double-check mechanism for classification. When computing the residuals of the ith class for classification, the collaborative representation mechanism considers not only the correlation between the associated representation and the intra-class representation, but also the correlation between intra-class representation and inter-class representation. Such double checking mechanism makes CRC more effective and robust for classification.

Limitations of CRC's residual model. SRC uses the value of representation residual norm over each class as criterion for classification. Zhang et al. [2] proposed an improving residual criterion by dividing the l_2 norm sparsity of representation coefficients as normalization. The two classification

* Corresponding author (email: maxiao2012@pku.edu.cn)
The authors declare that they have no conflict of interest.

residual criterions are quite similar as shown in

$$C^{'}(r_i) = (y - X_i\hat{\alpha}_i)^{\mathrm{T}}(y - X_i\hat{\alpha}_i), \tag{2}$$

$$C^{''}(r_i) = \frac{(y - X_i\hat{\alpha}_i)^{\mathrm{T}}(y - X_i\hat{\alpha}_i)}{\|\hat{\alpha}_i\|_2^2}. \tag{3}$$

When the training dictionary is overcomplete, the above Euclidean residual criterions are effective. The ith class's query sample y can be represented by $X_i\hat{\alpha}_i$ faithfully. The intra-class residuals tend to be white noise (here we suppose the noise component in face sample is a Gaussian-distributed white noise) and the inter-class residuals tend to be the query image itself as shown in Figure 1(a). The intra-class residual norms are significantly smaller than inter-class residual norms. The simple Euclidean criterions would result in good classification results in this case.

However, when the training samples are undercomplete, especially in undersampled cases with highly variable interferences caused by illumination misalignment, expression and occlusion, the Euclidean criterions may have some limitations. Suppose y is a query sample from class i, and the query sample has some types of variations which the ith class's training dictionary does not contain. If the jth class's training dictionary contains the same type of variations with y, y may be represented by $X_j\hat{\alpha}_j$ more faithfully than $X_i\hat{\alpha}_i$ in the collaborative representation mancinism. As shown in Figure 1(b), the intra-class residuals no more tend to be a simple white noise. Intra-class residual norms may be similar as or larger than inter-class residual norms. In these cases, it is not proper to use the Euclidean residual criterions.

The collaborative representation Bayesian recognition (CRB) method. The residual criterions for classification in SRC or CRC-RLS can be seen as an Euclidean similarity measure between y and the associated representation $X_i\hat{\alpha}_i$. By considering the information of whether the residual is characteristic of intra-class variations, a Bayesian residual model [3, 4] is exploited giving a novel probabilistic interpretation to the residual model in CRC. If y belongs to class i, residual $r_i = y - X_i\hat{\alpha}_i$ is characteristic of the intra-class variation Ω_{I}, where Ω_{I} corresponds to variations of the same individual, caused by illumination, poses and expressions. Then, the MAP Bayesian residual measure is defined as

$$C_{\mathrm{MAP}}(r_i) = P(\Omega_{\mathrm{I}}|r_i). \tag{4}$$

Suppose the intra-class residuals of Ω_{I} are Gaussian-distributed, an alternative ML Bayesian residual measure is defined as

$$P(r_i|\Omega_{\mathrm{I}}) = \frac{\mathrm{e}^{-\frac{1}{2}(y-X_i\hat{\alpha}_i)^{\mathrm{T}}\varSigma_{r_i}^{-1}(y-X_i\hat{\alpha}_i)}}{(2\pi)^{\frac{D}{2}}|\varSigma_{r_i}|^{\frac{1}{2}}}, \tag{5}$$

where D is the dimension of residual r_i, $\varSigma_{r_i}$ is the covariance matrix of the ith class's residual. Suppose all the intra-class residuals share the same residual covariance matrix $\varSigma_r = \varSigma_{r_i}, i \in C$. The Bayesian residual criterion can be simplified as

$$C_{\mathrm{ML}}(r_i) = (y - X_i\hat{\alpha}_i)^{\mathrm{T}}\varSigma_r^{-1}(y - X_i\hat{\alpha}_i). \tag{6}$$

However, the intra-class residual samples are quite difficult to capture. As face images are assumed to lie in a linear subspace [5–7], all the linear combinations of face images, including the reconstruction terms $X_i\hat{\alpha}_i$, also lie in the face linear subspace. If y belongs to class i, the intra-class representation residual $r_i = y - X_i\alpha_i$ can also be regarded as intra-class difference. In this article, we use the intra-class intensity differences from generic data to estimate $\varSigma_r$.

Combining the two models: the collaborative representation model for stable representation and the Bayesian residual model for robust classification, a novel collaborative representation Bayesian recognition (CRB) method is proposed. The detailed procedure of CRC is included in Appendix A.

The classification criterions in CRC are expressed as the ML forms of Bayesian residual criterions in CRB as

$$C'(r_i) = (y - X_i\hat{\alpha}_i)^{\mathrm{T}}(\lambda I)^{-1}(y - X_i\hat{\alpha}_i), \tag{7}$$

$$C''(r_i) = \left(\frac{y - X_i\hat{\alpha}_i}{\|\hat{\alpha}_i\|_2}\right)^{\mathrm{T}}(\lambda I)^{-1}\left(\frac{y - X_i\hat{\alpha}_i}{\|\hat{\alpha}_i\|_2}\right). \tag{8}$$

As the formulae above, the original collaborative representation based classification methods (including SRC and CRC-RLS) are the special cases of CRB when $\varSigma_r = \lambda I$, implying the items of intra-class residual vectors are independently and identically distributed. For criterions without normalization, the residuals are defined as $r_i' = y - X_i\hat{\alpha}_i$ $(i \in C)$. For criterions with normalization, the residuals are defined as $r_i'' = \frac{y - X_i\hat{\alpha}_i}{\|\hat{\alpha}_i\|_2}$ $(i \in C)$.

Back to the problem in the above, we will further discuss why CRB can achieve better performances in undersampled cases with highly variable intra-class variations. When the training dictionary is over-complete, the intra-class residual r_i tends to be a Gaussian-distributed white noise. So the independent-distributed assumption for intra-class residuals is reasonable in these cases. The

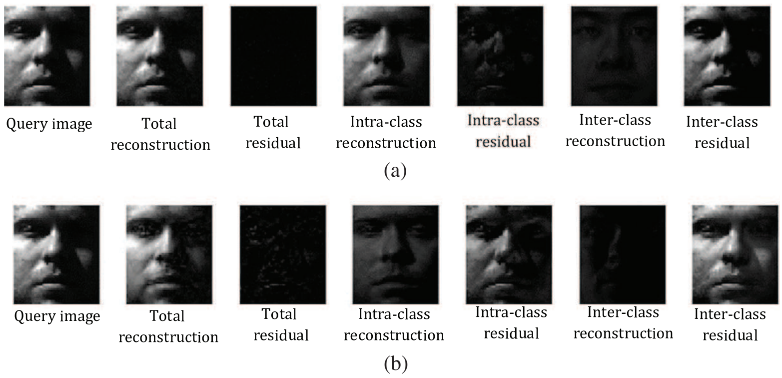

Figure 1 Reconstructions and residuals of collaborative representation based methods with over-completed (a) and under-completed dictionaries (b). From left to right: query image y, total reconstruction $X\hat{\alpha}$, total residual (reconstruction error) $y - X\hat{\alpha}$, intra-class reconstruction $X_i\hat{\alpha_i}$, intra-class residual $y - X_i\hat{\alpha_i}$, inter-class reconstruction $\sum_{j(j\neq i)} X_j\hat{\alpha_j}$, inter-class residual $y - \sum_{j(j\neq i)} X_j\hat{\alpha_j}$.

original collaborative representation methods can achieve similar (or even better) performances comparing to CRB.

While in the undersampled cases with highly variable intra-class variations, X_i is under-complete to reconstruct the query image. The intra-class residual r_i no more tends to be a Gaussian-distributed white noise as shown in Figure 1(b). As face images under varying illumination or pose variation approximately lie in a low dimensional linear subspace, items of high-dimensional residual vector are usually highly relevant and have different probabilistic distributions (Take faces images with varying poses as an example. Pixels lying on the rotation axis vary much more slightly compared to the pixels lying on the edge. There are correlations among them by the pose-transform model). The simple independent and identical distribution assumption on residual criterions is not quite reasonable in these cases. A more precise probabilistic model for describing the intra-class variations is needed. That is why the Bayesian residual model in CRB is more effective to distinguish the intra- and inter-class variations and make a better classification, especially in under-sampled cases.

Another advantage of CRB is its universality. For practical problems, the Bayesian residual model training process and face recognition process can be handled separately in two different databases. The universal Bayesian residual model is trained from large number of generic data, which makes full use of the prior knowledge contained in face samples. A series of experiments are conducted to demonstrate the efficacy of the proposed CRB method and validate the claims. The detailed experiment results are illustrated in Appendix B. The proposed CRB method can achieve about 5% improvement compared with the original collaborative representation based methods.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61333015) and National Basic Research Program of China (973) (Grant No. 2011CB302400).

Supporting information Appendixes A and B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Wright J, Yang A Y, Ganesh A. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell, 2009, 31: 210–227
2 Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of the International Conference on Computer Vision, Barcelona, 2011. 471–478
3 Pentland A, Moghaddam B, Nastar C. Bayesian face recognition using deformable intensity surfaces. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, 1996. 638–645
4 Moghaddam B, Jebara T, Pentland A. Bayesian face recognition. Pattern Recogn, 2000, 33: 1771–1782
5 Jacobs D W, Basri R. Lambertian reflectance and linear subspaces. IEEE Trans Pattern Anal Mach Intell, 2003, 25: 218–233
6 Qiao L S, Chen S C, Tan X Y. Sparsity preserving discriminant analysis for single training image face recognition. Pattern Recogn Lett, 2010, 31: 422–429
7 Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009. 1794–1801